UNIVERSITY OF DELHI
SCIENTIA ET MORES

RATAN TATA
LIBRARY

•

DELHI SCHOOL OF ECONOMICS

D.U.P. No. 1337—1-81—20,000

RATAN TATA LIBRARY
(Delhi University Library System)

Cl. No. B25 H6

Ac. No. 11915 Date of release for loan

This book should be returned on or before the date last stamped below. An overdue charge of Ten Paise will be charged for each day the book is kept overtime.

A CONCISE
MANUAL OF STATISTICS

A CONCISE MANUAL OF STATISTICS

(With Special Reference to the requirements of Students for Municipal Examinations)

by

CLEMENT BURTON

Associate of the Institute
of Municipal Treasurers and Accountants (Incorporated)
Fellow of the Royal Statistical Society

GEE AND COMPANY (PUBLISHERS) LIMITED
27-28 BASINGHALL STREET, LONDON, EC2

FIRST EDITION, 1937
SECOND EDITION, 1946

BOOK
PRODUCTION
WAR ECONOMY
STANDARD

This book is produced in complete conformity with the authorised economy standards

SET IN 10 PT. ON 11 PT. MONOTYPE MODERN NUMBER SEVEN

MADE AND PRINTED IN ENGLAND BY
GEE AND COMPANY (PUBLISHERS) LIMITED OF LONDON AND ST. ALBANS

CONTENTS

CHAP. I *Introduction* PAGE

Definition: Nature: Value: Limitations: Cautions: Statistical Methods: Treatment of Data: Functions of a Statistical Department 13

II *Collection of Data*

Preliminary Considerations: Sources: Methods of Collection: Questionnaires—General Considerations—Characteristics 19

III *Classification and Tabulation*

Objects: General Considerations: Rules for Tabulation: Aids to Clarity: Sampling—Attributes of a Sample—Methods: Frequency Tables—General Considerations—Cumulative Frequency Tables 22

IV *Accuracy and Approximation*

Accuracy : Error—Absolute—Relative—Percentage—Unbiased—Biased: Errors in Totals: Errors in Averages: Inexact Numbers—Addition—Subtraction—Division—Multiplication 27

V *Diagrammatic Presentation*

Pictograms—Advantages—Disadvantages: Circles: Rectangles: Bar Diagrams: Cartograms: Graphs—Characteristics—Construction—Logarithmic Curves: Frequency Curves—Construction—Normal Curve of Error—Smoothing—Histogram—Historigram—Skewness—Ogives: Lorenz Curve 33

✓VI *Averages*

Averages—Characteristics—Advantages—Disadvantages: Arithmetic Average — Advantages — Disadvantages : Weighted Arithmetic Average: Moving Average: Descriptive Average: Progressive Average: Geometric Mean—Advantages—Disadvantages: Harmonic Mean—Advantages—Disadvantages: Percentages: Ratios: The Mode—Advantages—Disadvantages: The Median—Advantages—Disadvantages : Quartiles : Deciles : Percentiles: Further Examples 62

CHAP. VII *Trend* PAGE

Trend: Types of Fluctuations: Investigation of the Trend and Short Period and other Fluctuations 79

VIII *Index Numbers*

Nature: Principles: Factors: Cautions in Use: Construction: General Principles: Fixed Base Method: Simple Index Numbers: Weighted Index Numbers: Chain Base Method: Published Index Numbers .. 89

✓ IX *Dispersion and Skewness*

Definition: Use of Graphic Methods: Absolute and Relative Measures: The Range: The Moments of the Dispersion: Average Deviation and Coefficient: Standard Deviation and Coefficient: The Modulus: Quartile Measure and Coefficient: Skewness—1st, 2nd and 3rd Measures and Coefficients: Further Examples 106

✓ X *Correlation*

Definition: Nature: Use of Graphic Methods: Coefficient of Correlation: Regression—Ratio of Variation: General Trend: Short Term Fluctuations: Time Lag 118

XI *The Normal Frequency Curve*

Probability: Introduction: Coin Tossing: Binomial Expansion: Normal Frequency Curve: General Description: Characteristics: Approximate Relationship of Measures of Deviation: Sampling: Law of Statistical Regularity: Law of Inertia of Large Numbers: Statistical Significance: Non-Measurable Characteristics: Measurable Characteristics: Distribution of the Means of Samples: Significance of the Difference between Means of Samples 129

XII *Municipal Statistics*

Object: Examples: For Information of Ratepayers—of Chief Officials—The Treasurer—The Medical Officer (including Estimates of Population)—and the Surveyor: As an aid to Supervision—The Treasurer—The Surveyor: In connection with Investigations—Examples 140

CHAP. XIII *Mechanical Aids to the Preparation of Statistical Data* PAGE

Advantages—Machinery—Adding—Listing—Calculating—Sorting and Tabulating: The Card System 157

XIV *Published Statistics*

Guide to Current Official Statistics—Official Statistics—Other Statistics 159

List of Symbols 175

Index 177

PREFACE TO THE SECOND EDITION

IN the 2nd Edition the concise nature and general layout of this work are preserved, but the scope has been extended generally by additions and insertions.

Further examples are introduced in several aspects of the subject, e.g. averages, dispersion, trend, &c., and extensions made in the pictorial illustration of data, especially by the use of graphic methods.

Amendments have also been made in certain cases, e.g. the description of published Index Numbers, where revisions have occurred since the date of publication of the 1st Edition.

A new chapter is included introducing the normal frequency curve and some of its implications.

May, 1946 C. B.

PREFACE

THE increasing importance of the subject of statistics in commercial and municipal activity has prompted the preparation of this handbook in the hope that it will prove of interest generally, and will be of assistance to those who are studying for the examinations of the various professional bodies of accountants, and more particularly for those of the Institute of Municipal Treasurers and Accountants (Incorporated).

The work, though concise, passes from elementary principles to the more advanced methods of the science in an order which, it is believed, will provide of their easy assimilation.

While many of the methods are applicable to data collected on a national basis, there is ample scope for the application of scientific methods of presenting data in all business houses and local authorities.

Various diagrams, &c., are given, illustrating the application of statistical methods and presentation of data, and chapters are included making suggestions as to the preparation of statistics in a municipality, and summarising the chief statistics published by government departments and others.

April, 1937 C. B.

Chapter I

INTRODUCTION

Definition: Nature: Value: Limitations: Cautions: Statistical Methods: Treatment of Data: Functions of a Statistical Department.

ORIGINALLY the term 'statistics' was applied to a branch of political science devoted to collecting and arranging facts relating to the resources of States in regard to wealth, power, extent, industry, population and similar matters. In the present specialised era the significance of the term has been narrowed considerably, and, apart from the popular description of a mass of data as 'statistics,' it may be defined as the science of analysis, comparison and deduction, involving the relationship of comparable, numerical facts.

The term is also used in the singular, when a 'statistic' refers to a statistical measure calculated from a representative sample of the data under review.

Statistics has also been termed the 'Science of Averages' and the 'Science of Estimates and Probabilities'. These phrases are partial truisms in that in the former case the principles of the science tend to lay emphasis on the general, rather than the particular characteristics of the data; and in the latter, owing to the fact that certain statistical methods are based on the mathematical theory of probabilities, when resort is made to estimation and approximation in dealing with the data under consideration.

The relationship of facts, rather than the facts themselves, is the subject of inquiry. Statistics are generally utilised for describing, analysing, classifying, and extracting salient features of large and complex masses of data and are not concerned with individual abnormalities or phenomena.

The value and uses of statistics lie not in the figures themselves, but in the deductions to be obtained from them. In some cases the object is the ascertainment of a definite fact, e.g. in connection with a proposed welfare service, it may be necessary to know the approximate number of children under 5 years of age, but in the vast majority of inquiries the object is to ascertain the trend or general tendency of particular aspects of the subject under review, e.g. the expansion or contraction of industries.

The position of the statistician lies between the accountant or cost accountant and the economist or executive authority.

An endeavour has been made in this work to treat the subject from the point of view of its relation to the operations of local authorities.

Many aspects of the science are thus of academic interest only but they are included, as familiarity with the essentials of the study is necessary in order that data may be correctly prepared and presented, also that the faults so commonly associated with the preparation and use of statistics may be avoided.

The chief values of statistical inquiry are to:—

(1) Provide information upon which estimates may be prepared, e.g. the probable yield of a tax; the possibilities of a business flourishing in a particular locality.

(2) Ascertain the causes of past events, i.e. whether they are—

- (*a*) accidental or of significance;
- (*b*) periodical or of a progressive nature;
- (*c*) exceptional or permanent,

e.g. in the calculation of life tables by insurance companies.

(3) Investigate the relations of comparable matters, e.g. the demand for goods and prices; the bank rate and bank reserve.

(4) Investigate effects, e.g. the maintenance of the gold standard in relation to the prices of commodities; the bank rate and its bearing on industry.

(5) Enable comparisons to be made of facts arising from time to time or from different sources, e.g. the rateable value of an area from creation to date, and as compared with different areas; the wages of workmen in different areas.

(6) Assist in financial investigations, e.g. the alteration of boundaries between local authorities; the purchase by a local authority of a privately-owned trading undertaking.

(7) Provide information as to processes, methods, &c., with a view to their improvement.

(8) Provide information as an aid to business control, supervision and audit.

(9) Present information with a view to assisting in courses of action or policy, e.g. adjustment of supply with demand; the necessity for the provision of a new school.

(10) Provide information to assist in forecasting probable future tendencies, e.g. the demand for gas or electricity.

(11) Provide information as to the trend of subjects of inquiry, e.g. the infant mortality rate.

Although the study of statistical data necessarily involves the scientific ascertainment of significant facts, or the relationship of facts, from a mass of material, the science has certain limitations. These limitations are:—

(1) The science of statistics, although presenting facts, is limited to

the numerical aspect only of the problems under review. Facts or tendencies deduced, therefore, must also be considered in conjunction with the relevant abstract factors, i.e. the social, economic, or political aspects of the case.

(2) In dealing with masses of data, individual peculiarities may become submerged. Often the trend of events only is required, but should these peculiarities be of material importance, deductions may prove unreliable.

(3) The subject of the inquiry may not be capable of numerical measurement, e.g. the poverty or health of the population. In such cases recourse must be made to material indicative of the factor under consideration, e.g. the number of people in receipt of public assistance or other relief, vital statistics, &c.

(4) In a case as indicated in item (3) it is highly desirable to obtain corroboratory evidence, as if an item used as an indicator is regarded as the sole factor, the statistical result may give incomplete information, or draw attention to part only of the question under investigation.

(5) Owing to practical difficulties in obtaining and presenting information, statistical results may prove to be estimates rather than exact statements, when account should be taken of the degree of accuracy in the data, and allowances made in order to offset omissions and errors.

Although a knowledge of the matter under consideration and statistical methods is presupposed when dealing with published statistics, great care is necessary in arriving at conclusions, owing to limitations or inaccuracies that may be present. The general cautions to be exercised in the use of statistics may be summarised as follows:—

1. *The Definition of Items.*
   The exact significance of the terms used should be apparent. The meaning of the term may be restricted for the particular inquiry, or, on the other hand, the scope of the term may prove to be too general for statistical purposes.
2. *The Homogeneity of Terms.*
   The significance of the term should not have changed during the period of the inquiry.
3. *The Relation of Data to the Object of Inquiry.*
   (1) Due allowance should be given to the object of the inquiry, its scope, and the real significance of the data collected.

   (2) The data should be truly representative of the subject under consideration.

   (*a*) It should be free from undue abnormalities or exceptional variations—unless the items are adequately numerous, or spread over a lengthy period of time, so as to ensure that any extremes are given their true significance as compared with the bulk of the data.

(*b*) Where it is not possible to consider the whole of the data *available, or certain items only are selected as the basis of investigation, it should be proved that the selected data* is in all respects representative of the subject under review.

(3) The data should be complete in all respects, or in any case complete for the purpose of the inquiry, e.g. a statement of the rate poundages in different areas, without a statement of the rateable values of comparable properties, may give an incorrect impression as to the actual cost of local services and administration.

(4) In the case of an inquiry where the data is not capable of numerical measurement, the suitableness and sensitiveness to changes of the indicator should be taken into account, and if possible, corroborative evidence obtained.

(5) Where a factor is taken as indicating a tendency, other evidence should be sought, e.g. 'exports per head' may not give a true index to prosperity, owing to the fact that one country may be self-supporting.

(6) The possibility of fallacious comparisons may arise, owing to the influence of factors, other than those under consideration, e.g. an increase in profits may not indicate prosperity where the increase has been caused by the inflation of the currency.

4. *The Status of the Compiler.*

The status and reliability of the compiler, e.g. government department, eminent statistician, &c., should also be considered, both from the point of view of the accuracy of the preparation and presentation of the material and that of the presence of personal bias.

5. *The Source of Information.*

Regard should also be had to the source of information, e.g. material supplied by local authorities to government departments is likely to be more reliable than that obtained from casual sources.

6. *The Period under Review.*

Unless short term influences are under consideration the data should, generally speaking, be obtained from information spread over a lengthy period.

7. *The Suitability of Averages.*
   Deductions from averages should not be greater than the figures warrant, as an average may, or may not, take into account extremes, or on the other hand, may be unduly affected by extremes, e.g. a local area may be considered wealthy as the rateable value per head of population amounts to £10. The result may be largely governed by the presence of a few extreme factors, the elimination of which would reveal a more accurate indication of the position. Also in considering averages the suitability or otherwise of the factors entering into the calculation require consideration, e.g. should the number of cases of diphtheria be considered in relation to the total population or to the population who have been immunised?
8. *The Degree of Accuracy.*
   The degree of accuracy of the data must also be taken into account in considering statistical matter, and adequate allowance made in deducing conclusions.
9. *Technical &c., Errors.*
   The possibility of technical errors may arise in the preparation of data, e.g. taking an average of percentages; or misleading conclusions may be induced by badly drafted diagrams.
10. *Basis of Comparison.*
    The method adopted of dealing with statistical matter may not be suitable, e.g. a reduction of 10 per cent. in two items (i.e. the relative reduction) may indicate equal results, when the absolute reduction, say 100 and 125,000, may be a more suitable basis or vice versa.
11. *Bias.*
    The investigation should be conducted without personal bias and with an unprejudiced mind as to the significance of, and the ultimate conclusions to be deducted from, the data.

Statistical methods are mathematical, pictorial, tabular, graphic, &c., devices whereby the relation of facts is made apparent. The term also applies to agreed methods of the presentation and mathematical treatment of data. The term 'applied statistics' refers to the practical application of statistical methods.

As will be indicated later, each method of presentation, each average, &c., has its limitations, advantages and disadvantages, consequently a knowledge of statistical methods is necessary to the correct presentation and appreciation of statistical data, and the avoidance of erroneous conclusions.

The preparation of statistical data involves the following operations:—

(1) The collection of the data (e.g. by the issue of questionnaires).

(2) The tabulation of the data (e.g. by constructing a frequency table).

(3) The summarisation of the data by one or more of the following methods:—

- (*a*) Constructing a pictogram (i.e. the diagrammatic presentation of numerical facts).
- (*b*) Constructing a graph (i.e. the diagrammatic presentation of numerical facts).
- (*c*) Taking a sample (i.e. a representative item or series of items).
- (*d*) Calculating an average (i.e. a representative item).
- (*e*) Arising from the foregoing the following calculations may be necessary:—
  - (1) The preparation of an index number.
  - (2) The measurement of the variation of individual items from a selected sample or an average (termed 'dispersion').
  - (3) The measurement of the numerical relationship existing between sets of facts, by calculating a 'coefficient of correlation'.
  - (4) The investigation of the trend of the data and the long and short term fluctuations to which it is subject.

The functions of the statistical department of a local authority would be:—

(1) To collect and generally prepare by tabulating, graphing, &c., to file, index and investigate and report as to the significance of such statistics as may be deemed to be of importance to any department or committee of the local authority, or may be required by the chief officers of the various departments.

(2) To correlate the data relating to the area with that of other comparable areas.

(3) To make suggestions as to the extension of statistical services from time to time with a view to more efficient administration.

Chapter II

COLLECTION OF DATA

Preliminary Considerations: Sources: Methods of Collection: Questionnaires—General Considerations—Characteristics.

Before the collection of data is undertaken the undermentioned factors should be considered, viz.:—

1. The scope of the inquiry in regard to the sources of the information and the extent to which it is desired to investigate the subject.
2. The determination of the information required.
3. The exact definition of the form in which the information is to be presented.

In the case of many municipal inquiries, much of the information in greater or less detail will be available. It may thus be possible to classify the data and make a preliminary investigation prior to seeking more detailed information by, for example, the issue of questionnaires.

The main sources of statistical data are:—

(1) Official tables, &c., published by government departments—
  (*a*) Periodical reports, e.g. Local Government Financial Statistics, the Census.
  (*b*) Special reports of inquiries by royal commissions, e.g. the Royal Commission on Local Government.

(2) Tables, &c., published by local authorities, e.g. annual reports, abstracts and epitomes of accounts.

(3) Reports and tables published by trades associations, chambers of commerce and trade supplements of newspapers, e.g. *The Times Trade Supplement.*

(4) Reports and journals published by professional associations and universities, e.g. the Journal of the Royal Statistical Society.

(5) Reports, &c., published by private individuals and firms, e.g. The Preston Rates Return.

(6) Private firms and individuals.

Data may be collected:—

(1) Personally.

(2) By enumerators who interview the informants and, by the information so derived, complete the schedules of points under consideration.

(3) From accredited correspondents, e.g. crop reports, meteorological and ornithological investigations.

(4) By the issue of questionnaires.

(5) From data prepared and issued by government departments, local authorities, trade and professional associations and technical journals.

The relative merits, demerits and cautions in the use of questionnaires apply in greater or less degree to each of the methods set out above. In this connection the following factors arise:—

(1) Terms used require exact definition, e.g. 'working week', 'weekly wage', 'profit', 'employee' . . . are capable of many interpretations.

(2) The homogeneity of the items. The terms used must be of the same significance throughout the inquiry, e.g. a comparison of rate poundages may lead to fallacious deductions unless the basis of assessment is the same in all cases. The terms ship, gun, motor-car, have a different aspect at the present time from that of a few years ago.

(3) Facts, and not opinions, should be sought, and as far as possible personal bias avoided in an endeavour to obtain impartial data. More reliance can be placed on information obtained from a government department or other impartial source than from private individuals who may have personal interests in the results accruing from the investigation. Generally speaking, official sources are infinitely more reliable in this respect, as often there is statutory power to demand, and verify by audit, the information required.

(4) Owing to the possibility of fallacies, or technical errors, arising in the calculation or preparation of the data, the information required should be the actual data, and not figures deduced therefrom.

A well drafted questionnaire should have the under-mentioned characteristics:—

(1) The questions should not be ambiguous or admit of differing interpretations.

(2) The questions should be so worded as to require a simple answer, or figures, free from personal bias.

(3) Where an opinion is required, the questions should be so framed as to require an unqualified affirmative or negative.

(4) The questions should be framed so as to be corroboratory.

(5) Unnecessary information should not be solicited, nor should questionnaires be too inquisitorial.

(6) Should the completion of the form be voluntary, a statement as to the reason for the questionnaire and the use to which it is to be put will often obtain the co-operation of the person to whom it is addressed.

(7) The solicitation of confidential information should, if possible, be avoided.

(8) Where necessary, the terms used should be carefully defined.

(9) The answers to the questions should admit of their subsequent tabulation.

(10) Other aspects being equal, the questions in a well drafted questionnaire should be—

(*a*) short and easily understood;
(*b*) simple to answer;
(*c*) few in number.

Chapter III

CLASSIFICATION AND TABULATION

Objects: General Considerations: Rules for Tabulation: Aids to Clarity: Sampling—Attributes of a Sample—Methods: Frequency Tables—General Considerations—Cumulative Frequency Tables.

A MASS of data, as such, is of very little, if any, use until it has been tabulated and classified.

All phenomena have a tendency to cluster round certain types or to exhibit characteristics which enable it to be placed into certain classes, groups or series: thus the method of preliminary classification is of the utmost importance as, unless it is carefully carried out, the real import of the inquiry may be obscured, incorrect deductions obtained, or even the collection of the data rendered abortive.

The object of statistical analysis is to ascertain—

(1) The variations existing between the individual items of the data.
(2) The variations existing between groups of data as a whole.

Statistical tables may be classified under two general headings:—

(1) *Original*, where the data is presented in the form in which it was collected.
(2) *Derivative*, where the data has been subjected to processes such as classifying, grouping, totalling or averaging.

The items may be—

(1) classified according to numerical strength, size, &c.;
(2) arrayed in ascending or descending order;
(3) tabulated so as to render their defined or variable characteristics apparent,

according to the particular object for which the work is undertaken.

The object of the inquiry may be to obtain—

(1) descriptive data, when the material is dealt with from the historic standpoint, or
(2) scientific data, when the material is dealt with so as to provide material to verify or disprove some hypothesis or to establish an economic law,

but whatever method of presenting the facts may be decided upon, the statistical tables should be accompanied by notes or a statement stating—

(1) definitions of items;
(2) the sources from which the information was derived;

(3) the basis of the calculations made;
(4) the probable sources of error, i.e. the approximations made, and, if desired,
(5) notes drawing attention to salient facts, or
(6) the deductions obtained.

The considerations arising when classifying information may be summarised as follows:—

(1) Tables may be either—
    (*a*) simple, i.e. showing one set of information only, e.g. the number of workmen employed by a local authority, over a number of years;
    (*b*) complex, i.e. showing several aspects of the subject under review, or giving the analysis of totals, e.g. the number of workmen employed by a local authority, over a number of years, analysed under the work on which they were engaged.
(2) Tables should be so devised as to answer the question implied.
(3) The title to the table or tables should be self-explanatory, and not admit of ambiguity.
(4) Although a brief heading is preferable to a lengthy statement, the latter, if concise and well worded, is of infinitely greater value than a heading that admits of several meanings.
(5) The size of the table(s) should be kept within reasonable limits, as masses of figures and large tables tend to confusion. Folded tables are undesirable.
(6) Generally speaking, one table showing the whole of the information is first prepared, supported by subsidiary tables illustrating particular aspects of the matter under review, or giving greater detail.
(7) If permissible, figures should be presented as simply as possible, e.g. given to the nearest 1,000, as large numbers tend to destroy the legibility of the table as a whole. Where abbreviation of figures is adopted, the last digit should be accurate.
(8) Where items require special explanation, footnotes may be added. Should these notes be numerous, they tend to confusion. In such cases the preparation of a subsidiary table may be of value.
(9) The items to be compared should in themselves be comparable.

Points also arise in the actual setting out of the table, the object being to give prominence to important figures, and to admit of their easy assimilation, e.g.:—

(1) The importance of headings and sub-headings should be indicated by suitable type.
(2) The thickness of ruling and width of columns may be arranged to assist in distinguishing important from subsidiary matter.

(3) Each main heading should be supported by figures giving the 'totals', 'averages' or 'percentages'. These should be placed in proximity to the figures from which they are derived.

(4) Figures to be compared should be adjacent, or as near to one another as practicable.

(5) Data should be arranged in blocks, i.e. a small number of lines (say 5), a space, and a further block of data. This is important where the width of the table is considerable. The repetition of the headings or the numbers of the lines on either side of the table may assist in this direction.

(6) Data tabulated vertically is preferable to the horizontal treatment.

(7) The placing of unimportant data in a miscellaneous column may considerably reduce the number of columns required.

The volume of data may be so great that its tabulation and classification could only be completed at the expense of considerable time and labour, perhaps out of all proportion to the value of the work when performed; it may be impossible to obtain complete information, or sufficiently reliable information could only be obtained from representative groups of material. In these cases 'sampling' would be permissible. 'Sampling' is the term applied to selecting representative facts from a large mass of data. The use of this method is widely adopted in business. Contracts for the supply of goods are often based on the submission of samples, while auditors make 'spot' or 'test' checks of the books under audit.

Note. The terms applied to the data as a whole from which the sample is drawn are: the 'population' or the 'universe'.

The scientific basis of sampling, and, in fact, the majority of the statistical methods, is the 'Law of Statistical Regularity' which is derived from the mathematical theory of probabilities. The Law of Statistical Regularity states that a representative number of items, selected at random from a larger group is, on the average, reasonably certain to contain the characteristics of the larger group.

A corollary of the Law of Statistical Regularity known as the 'Inertia of Large Numbers' states that in large masses of data abnormalities will occur, but in all probability phenomena in one direction will be offset by phenomena in the opposite direction, and the average will remain constant.

It is, therefore, necessary that where sampling is adopted, to ensure that the number of samples taken and the period of review are adequate, so that the result of the inquiry may be accurate.

The attributes of a good sample may be summarised as follows:—

(1) Samples should be truly representative of the subject under review.

(2) Personal or other bias in the selection of samples should be avoided.

(3) Every item must have an equal possibility of being selected as a sample.

(4) Accurate grouping of the data is a necessary preliminary in order to diminish the possibility of the selection of unrepresentative samples.

(5) The tendency to select the most easily accessible or obvious items should be avoided.

(6) The larger the mass of data and number of samples selected, the more accurate will be the result.

The chief methods of selecting samples are:—

(1) *By Lottery.*—It may not be possible to obtain information from all sources. These scources could be given numbers and the actual people to be approached obtained by lottery.

(2) *By Arraying.*—The data could be arrayed in groups, or in ascending or descending order, and every 5th, 10th, 20th, &c., item taken.

(3) *By Grouping.*—Groups could be sub-divided and certain typical examples examined in detail.

(4) *By Personal Selection.*—Typical samples could be obtained by an impartial selection.

In cases of sampling it is desirable that two or more methods are adopted so that results can be compared and, if necessary, better methods devised, or where personal bias might arise, the same method performed by more than one person.

Frequency Tables.—If masses of data are at all large, it is usually necessary to classify and sub-divide it into classes and groups. There are no fixed definitions of these terms, but it is generally recognised that a class signifies that the items under consideration have a distinguishing feature which is not capable of measurement, and a group a further sub-division of the data, the basis of which is some distinguishing feature capable of measurement, e.g. road stone could be divided into the classes of granite, Kent ragstone, limestone, &c., and these classes into groups under the heading of price, relative hardness, or other numerical features. Investigations of matter falling into classes may be termed the 'statistics of attributes' and of items falling into groups the 'statistics of variables'.

The next step is to array the data, i.e. to arrange the items in the order of their magnitudes. On classifying and arraying the data, and introducing some measurable characteristic, e.g. time, a statistical series is formed, i.e. each item is numerically compared under the heading of certain measurable characteristics.

The term 'continuous' is applied to a series where the data, which is not capable of exact measurement, where exact measurements are not available, or where items pass from value to value with small differences, falls within certain class limits, e.g. the ages of a large number of school children.

A 'discrete' or broken series results where the data is capable of exact measurement, e.g. the rateable values of properties.

After classification into classes and groups and arraying, a frequency table may be constructed. A frequency table is a tabular statement of data, suitably sub-divided and showing, in ascending or descending order, the number of items (the frequency) in each sub-division. (See Example 1.)

The following steps are necessary in the preparation of a frequency table:—

(1) To determine the size of the 'class' or 'group' intervals. Class-intervals should be of equal size throughout in order to facilitate comparison. For similar reasons class-intervals should not be unduly numerous.

(2) To determine the limits of the intervals, e.g. the limits of the intervals or the mid-values may be round numbers, 50-100, 100–150, &c., or .5–1.5, 1.5–2.5, &c.

For example, it may be necessary to prepare an analysis of the properties appearing in the valuation list. These could be placed in the classes of dwelling-houses, shops, garages, industrial undertakings, and grouped under rateable value. An example of a frequency table, with a column showing the cumulative number of dwelling-houses under the heading of rateable value, is given below. (Example 1).

A cumulative frequency table is the table resulting when class frequencies, instead of being shown separately, are totalled so as to show the total frequency at different parts of the table. (See Example 1.)

Example 1. DWELLING HOUSES

Rateable Value	Frequency	Cumulative Frequency
Not Exceeding £10	961	961
Exceeding £10 and not exceeding £20 ..	2,460	3,421
,, 20 ,, ,, 30 ..	5,712	9,133
,, 30 ,, ,, 40 ..	1,503	10,636
,, 40 ,, ,, 50 ..	546	11,182
,, 50 ,, ,, 60 ..	95	11,277
,, 60 ,, ,, 70 ..	51	11,328
,, 70 ,, ,, 80 ..	22	11,350
,, 80 ,, ,, 90 ..	5	11,355
	11,355	

Chapter IV

ACCURACY AND APPROXIMATION

Accuracy: Error—Absolute—Relative—Percentage—Unbiased—Biased: Errors in Totals: Errors in Averages: Inexact Numbers—Addition—Subtraction—Division—Multiplication.

Even though details of items under consideration may be given exactly, which in many cases is improbable, it may be necessary for purposes of calculation or presentation, to present them shorn of digits, or otherwise approximated. The object of the inquiry may be to obtain the 'trend' of particular events, i.e. the general indications, in distinction to peculiar fluctuations, when extreme accuracy would be superfluous. Again, spurious accuracy may be both unnecessary and confusing. A railway company may employ 144,362 men. For most purposes it would only be necessary to state 145,000 or 144,500. The degree of accuracy to give dependable results will vary according to the

(1) magnitude of the items, e.g. the yearly imports of wheat may be approximated to the nearest 100,000 bushels;

(2) the intrinsic value of the items, e.g. iron as compared with platinum.

If approximation is adopted, a note as to the degree of 'error' appearing in the facts should be given. The term 'error' in statistical terminology does not mean a mistake, but refers to the difference or estimated difference between the approximated and the actual data.

The degree of 'error' arising in statistical data may be measured either absolutely or relatively. For example, 10,675 may be contracted to 10,700 when the:—

'absolute error', i.e. the difference between the actual figures and the estimate, would be 10,700—10,675=25, and the 'relative error', i.e. the ratio of the absolute error to the estimate

$$\frac{25 \quad \text{(the absolute error)}}{10{,}700 \quad \text{(the estimate)}} = .002336$$

The 'percentage error' is .2336, i.e. the 'relative error' × 100.

The relative error forms a useful coefficient for comparing the degree of error of different sets of data.

Errors are also divided into two classes, i.e.—

Unbiased (compensating or fortuitous) and
Biased (cumulative).

Unbiased Errors, are those where the estimated or contracted figures may be on either side of the actual figures, i.e. taken to the nearest round figure, e.g. 956 and 1,041 would each be recorded as 1,000. Unbiased errors tend to eliminate each other in a total, and, should the frequency of the items be large, the result would not be inaccurate to any material degree. Thus the size of the error (unbiased) varies inversely in proportion to the number of items.

Example 2.

(The size of the table is necessarily contracted.)

Actual Figures	Estimated Figures
56	100
141	100
248	200
1,072	1,100
184	200
194	200
1,895	1,900

Biased Errors, are those where the estimated or contracted figures are all on the same side of the actual figures, e.g. 906 and 996 would each be recorded as 1,000 or 900.

Owing to their nature, biased errors assume magnitude directly in relation to the number of items.

Example 3.

Actual Figures	Estimated Figures (nearest 100 above)	Estimated Figures (nearest 100 below)
56	100	—
141	200	100
248	300	200
1,072	1,100	1,000
184	200	100
194	200	100
1,895	2,100	1,500

Error in Totals.—Often in the collection of data the degree of error is known but the actual amount of the error unknown. The calculations necessary to estimate the amount of error are as follows:—

1. *Unbiased Errors.*—When figures are taken, for example, to the nearest 100, the maximum possible error is 49; but as the error may be anywhere between 0 and 49 the average or probable error is 24.5.
   - (*a*) To estimate the absolute error in the sum of a number of items containing unbiased errors, multiply the probable error by the square root of the number of items. (Example 4, line 3, column 2.)
   - (*b*) To estimate the relative error in the sum of a number of items containing unbiased errors, multiply the probable error by the square root of the number of items and divide the result by the estimated total. (Example 4, line 4, column 2.)
2. *Biased Errors.*—When figures are taken, for example, to the nearest 100 above, the maximum possible error is 99, but on the average, the most probable error would be 49.5.
   - (*a*) To estimate the absolute error in the sum of a number of items containing biased errors, multiply the probable error by the number of items. (Example 4, line 3, column 3.)
   - (*b*) To estimate the relative error in the sum of a number of items containing biased errors, multiply the probable error by the number of items and divide the result by the estimated total. (Example 4, line 4, column 3.)

Errors in Averages.

1. *Unbiased errors.*
   - (*a*) To estimate the absolute error of an average, divide the probable error of the items by the square root of the number of items. (Example 4, line 7, column 2.)
   - (*b*) To estimate the relative error of an average, divide the probable error of the items by the square root of the number of items and also by the average of the items. (Example 4, line 8, column 2.)
2. *Biased Errors.*—Biased errors, owing to their nature, remain unaltered in an average.
   - (*a*) The probable absolute error in an average is thus the probable absolute error of any particular item. (Example 4, line 7, column 3.)
   - (*b*) The probable relative error is the probable absolute error of any particular item divided by the average of the items. (Example 4, line 8, column 3.)

An example showing the application of these methods is given below, column 1 giving the actual, and columns 2 and 3 the estimated figures. The errors in column 2 are unbiased and in column 3, biased.

Example 4.

		Actual Figures (1)		Approximated Figures to nearest hundred (2)		Approximated Figures to nearest hundred below (3)
		253		3		2
		159		2		1
		443		4		4
		191		2		1
		576		6		5
		334		3		3
		198		2		1
		247		2		2
		703		7		7
1. Totals		3,104		3,100		2,600
2. Actual absolute error				4		504
3. Probable absolute error			$(25\times\sqrt{9})$	75	(50×9)	450
4. Relative error			$(75\div3{,}100)$	.024	$(450\div2{,}600)$	.173
5. Average	$(3{,}104\div9)$	345	$(3{,}100\div9)$	344	$(2{,}600\div9)$	289
6. Average error			$(4\div9)$	.4	$(504\div9)$	56
7. Average probable error			$(25\div\sqrt{9})$	8.3		50
8. Average relative error			$\dfrac{25}{\sqrt{9}\times344}$	.024	$(50\div289)$	.173

It will be observed that the probable absolute error of the averages (8.3 and 50) is equal to the probable absolute error of the totals divided by the number of items ($75\div9=8.3$ and $450\div9=50$).

Also that the relative error of the average is equal to the relative error of the total (.024 and .173).

Further, that the average probable error is less than the probable absolute error of any individual item in the case of unbiased errors (8.3 and 25) and is the same (50) in the case of biased errors.

Note.—The term 'probable' is used to indicate the average of any errors that may arise, and not the particular error which may occur most frequently. The occurrence of any error is equally likely, and in some cases, for example in a rectangular distribution, no particular error predominates.

Many examples of the mathematical treatment of estimated or inexact figures could be given, the actual arithmetic necessary varying with the problem under consideration. The undermentioned examples of addition, subtraction, division and multiplication are given as illustrations.

Example 5.

Addition.—The sum of groups totalling 8,000, 4,000, 86,000 and 270,000 correct to 5 per cent., 1 per cent., .5 per cent. and .25 per cent. respectively, is required.

$$\begin{array}{rcr}
8{,}000 & \pm & 400 \\
4{,}000 & \pm & 40 \\
86{,}000 & \pm & 430 \\
270{,}000 & \pm & 675 \\
\hline
368{,}000 & \pm & 1{,}545 \\
\hline\hline
\end{array}$$

The result is $368{,}000 \pm 1{,}545$ or between 369,545 and 366,455.

Example 6.

Subtraction.—45,500 is to be subtracted from 125,000. The figures are correct to 3 per cent. and .5 per cent. respectively.

Maximum difference	Minimum difference
125,625	124,375
− 44,135	− 46,865
81,490	77,510

The result is $79{,}500 \pm 1{,}990$ or between 81,490 and 77,510.

Example 7.

Division.—It is required to divide 171,600 (correct to 1 per cent.) by 1,430 (correct to 5 per cent.)

Formula.—Let a = dividend
b = divisor
x = possible error of dividend
y = possible error of divisor.

Then the answer lies between $\dfrac{a + x}{b - y}$ and $\dfrac{a - x}{b + y}$.

Applying the formula—

$$\frac{171{,}600 + 1{,}716}{1{,}430 - 71.5} = 128, \quad \frac{171{,}600 - 1{,}716}{1{,}430 + 71.5} = 113$$

the result is—

120.5 ± 7.5 or between 113 and 128.

Example 8.

Multiplication.—It is required to multiply 450 (correct to 2 per cent.) by 2,600 (correct to .5 per cent.)

Formula.—Let m = multiplier
n = multiplicand
x = possible error of multiplier
y = possible error of multiplicand.

$$(m+x)\quad(n+y)=mn+nx+my+xy$$
$$(m-x)\quad(n-y)=mn-nx-my+xy)$$
$$=(mn+xy)\pm(nx+my)$$

Note.—In practice the value of xy, compared with the actual figures, is so small as to be of no material importance.

Eliminating xy the formula then becomes—

$$mn\pm(my+nx)$$

Applying the formula—

$$450\times2{,}600=1{,}170{,}000\pm((450\times13)+(2{,}600\times9))$$

The result is 1,170,000±29,250 or between 1,199,250 and 1,140,750.

Chapter V

DIAGRAMMATIC PRESENTATION

Pictograms—Advantages—Disadvantages: Circles: Rectangles: Bar Diagrams: Cartograms: Graphs—Characteristics—Construction—Logarithmic Curves: Frequency Curves—Construction—Normal Curve of Error—Smoothing—Histogram—Historigram—Skewness—Ogives: Lorenz Curve.

Pictograms.—The term‘ pictogram’ is applied to diagrams of a pictorial or semi-pictorial nature, designed to draw attention to statistical matter. The value of pictograms is restricted from the point of view of the statistician, subject, however, to the use of graphs, but they form a useful and popular method of drawing attention to facts and salient features of data. For example, an epitome of accounts embodying a well designed set of pictograms is more likely to evoke the interest of ratepayers in the affairs of a local authority than a booklet composed of sets of tables; although the latter would, generally speaking, contain the most information.

It adds greatly to the value of pictograms if they are supported by the actual figures from which they are derived.

The advantages and disadvantages of pictograms are summarised below.

Advantages.

1. Pictograms form an excellent medium for drawing attention to salient facts; and for showing the relation of parts to each other, and to the whole of the data.
2. They may suggest the direction for further investigation.
3. Pictograms are easily comprehended.
4. Diagrams, compared with tables of figures, convey a more vivid impression.
5. Pictorial methods of presentation will probably evoke interest, where tabulated data might be ignored.

Disadvantages.

1. Pictograms merely illustrate data and from the point of view of a skilled observer, do not add anything to its value, subject to the use of the graphic method (*post*).
2. They are only suitable for presenting simple facts.
3. Most diagrams are apt to become confused and defeat their object if an endeavour is made to present several aspects of the data under review on the same diagram.

4. Approximated data only can be presented.
5. Misleading conclusions can easily be deduced unless great care is exercised in drafting the diagram, e.g. the scales, &c., of the diagram readily admit of the exaggeration or diminution of facts.
6. Certain forms of pictograms, although geometrically correct, are apt to be misleading, e.g. diagrams such as circles and rectangles involving the comparison of areas.

General suggestions as to the pictorial presentation of data are given below. They are, however, in the nature of general rules, the accurate and clear presentation of the data being the primary consideration:—

1. The arrangement, generally, of the diagram should be from left to right.
2. Lineal magnitudes are a better form of presentation, than the use of areas or volumes.
3. The lines of the diagram should be boldly distinguished from the ruling.

Other points in regard to the preparation of line charts are given under the heading of 'Graphs'.

Examples of Pictograms.

1. *Circles* (or Pie charts)
   (*a*) Circles divided into sections to show the aggregate and its parts. The area of the circle represents the aggregate of the data and the individual items are shown by segments.
   Figure 1.
   (*b*) Separate circles whose areas or diameters are proportional to the items to be compared.
   Figure 2.
2. *Squares.*
   Separate squares whose areas are proportional to the items to be compared.
   Figure 3.
3. *Rectangles or Block Diagrams.*
   (*a*) Separate rectangles whose areas are proportional to the items to be compared.
   Figure 4.
   (*b*) Rectangles divided into sections to show a total and its parts. The area of the rectangle represents the aggregate of the data and the individual items are shown by the sections.
   Figure 5.
4. *Bar Diagrams.*
   The size of the items are shown by the length of the bars. Bars may be sub-divided to show the component parts of the items

Figure 1.

RATES LEVIED 1935-36.

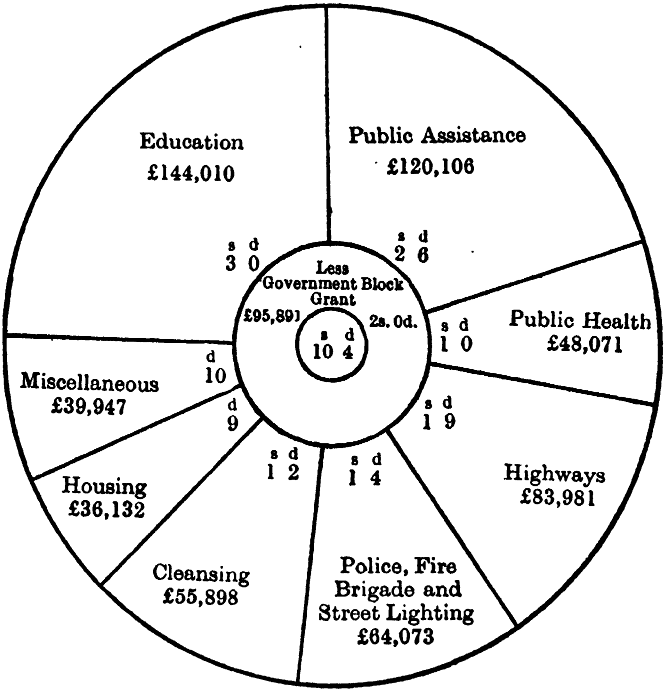
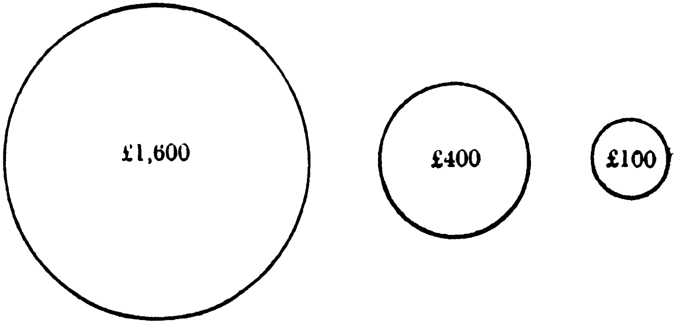


Figure 2.

DIAGRAMS SHOWING COMPARISONS OF AREAS — CIRCLES.



similar to the rectangle illustrated in Figure 5. This type of diagram admits of the comparison of more than one factor. Figures 6 and 7 illustrate the introduction of the time element as well as different subjects.

Figure 3.

DIAGRAMS SHOWING COMPARISONS OF AREAS — SQUARES.

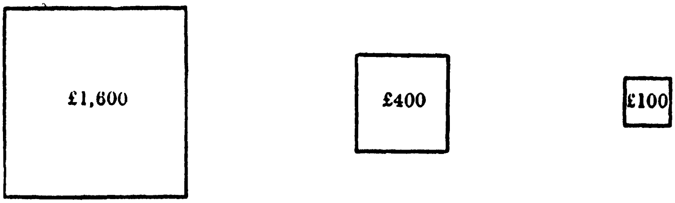
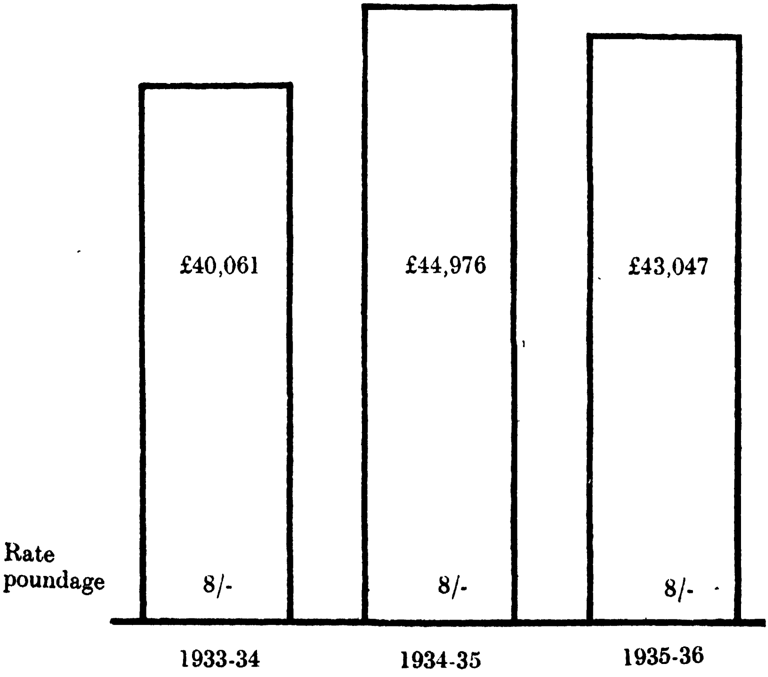


Figure 4.

RECTANGLES TO SHOW PRODUCTS

RATES COLLECTED

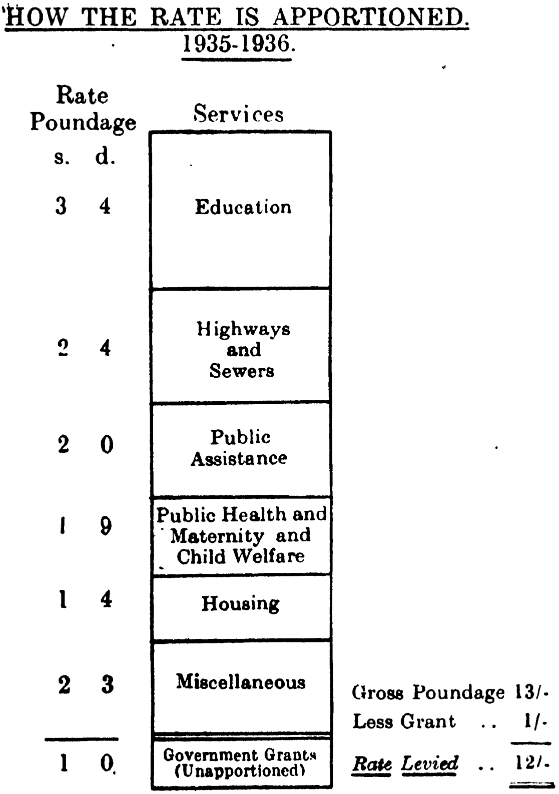


5. *Cartograms.*

Pictograms are termed cartograms where recourse has been had to shading, hatchuring, &c., to distinguish features, e.g. the population of different areas may be shown by appropriately shading or colouring a map of the country.

Figure 5.

RECTANGLE TO SHOW A TOTAL AND ITS PARTS.



Note.—1. Items are shown net, i.e. less specific income.

2. Government Grants under the Local Government Act, 1929, are in aid of expenses generally and are thus deducted from the total rate poundage.

Figure 6.

BAR DIAGRAM.

COMPARISON OF NET LOAN DEBT and PERMANENT ASSETS per Head of Population.

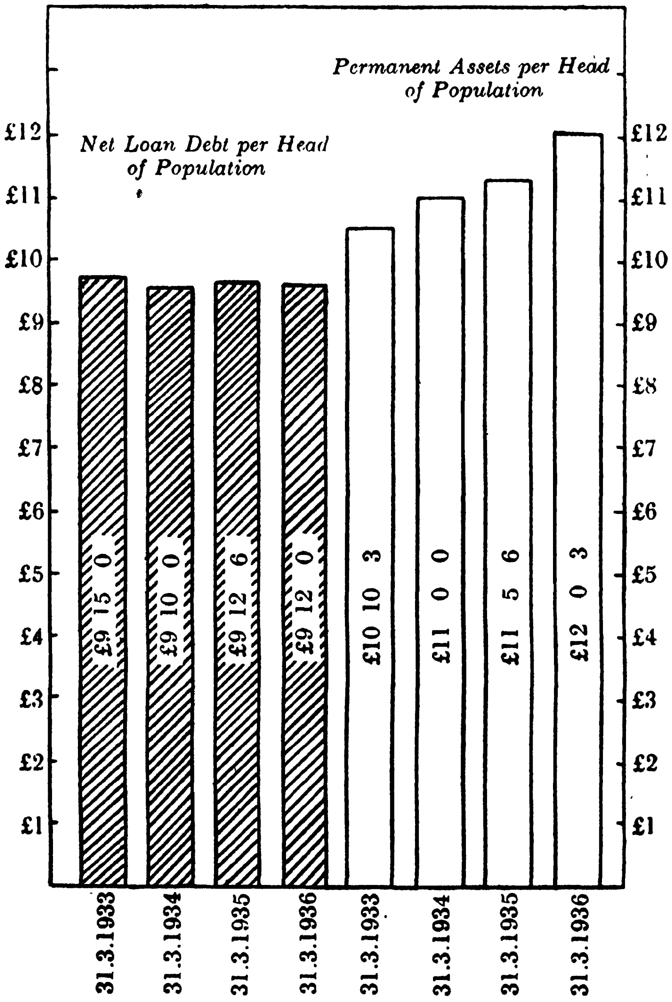


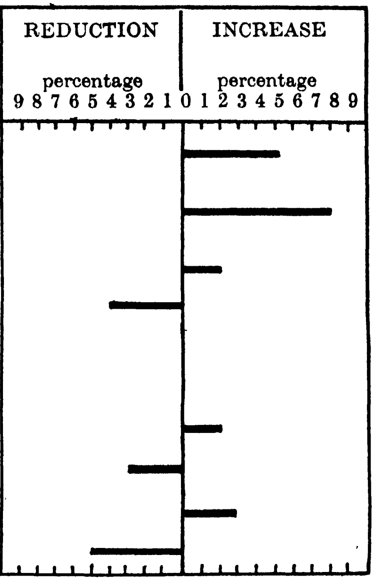
Figure 7.

BAR DIAGRAM SHOWING

PERCENTAGE REDUCTION OR INCREASE OF EXPENDITURE

1935-6, 1936-7.

SERVICE	1935/6 £ 000	1936/7 £ 000	Percentage −	Percentage +
Education ..	144	151		5
Public Assistance ..	120	130		8
Public Health	48	49		2
Highways ..	84	81	4	
Police, Fire Brigade and Street Lighting	64	66		2
Cleansing ..	56	54	3	
Housing	36	37		3
Miscellaneous..	40	38	5	



6. *Graphs* or *Line Charts* form another convenient medium for the symbolical expression of numerical facts. Essentially they are a method of presenting data of a two-dimensional nature by a system of rectangular co-ordinates. The uses of graphic methods in presenting data vary from their use as pictograms, i.e. in illustrating simple facts, to their use in the most advanced stages of the science, e.g. as an aid to investigating complex relationships. They are exceptionally valuable when investigating the statistics of variables.

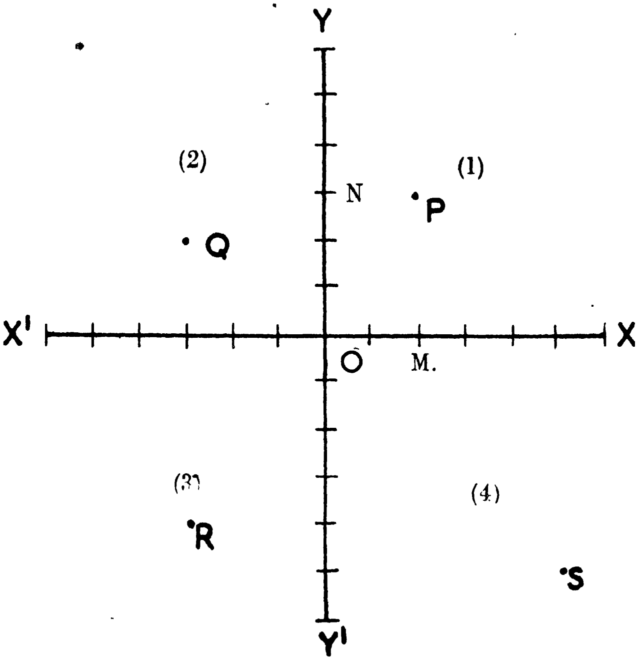
The characteristics of a graph are as follows:—

(1) A graph consists of two straight lines intersecting at right angles, known as the axes of reference, one horizontal—the x axis or axis of abscissae, and the other vertical—the y axis or axis of ordinates.

(2) The x and y axes divide the plane into four quadrants, numbered 1, 2, 3 and 4 respectively.

Figure 8.

GRAPH OR LINE CHART



Notes.—

The *x* axis	XOX^1
The *y* axis	YOY^1
The Quadrants	The planes (1), (2), (3) & (4)
The point of origin	O
Positive values of *x*	OX
Positive values of *y*	OY
Negative values of *x*	OX^1
Negative values of *y*	OY^1
Abscissa of point P	OM
Ordinate of point P	ON

The position of the points (2, 3), (—3, 2), (—3, —4) and (5, —5), alternatively set out as follows:—

x	*y*
+2	+3
—3	+2
—3	—4
+5	—5

are indicated by the letters P, Q, R and S.

(3) The point of intersection of the two lines, i.e. the x and y axes, is termed the origin or zero point.

(4) Distances measured from the origin upward and to the right are regarded as positive, and downward and to the left, negative.

(5) The x and y axes are divided into suitable scales, not necessarily equal, but chosen to suit the convenience in illustrating movements or other features of the data under review.

(6) The value of each pair of variables is represented by a point, the position of the point being determined by the value of the variable in terms of the x and y scales. These values are known respectively as the abscissa and ordinate of the point. The abscissa and ordinate of a point are collectively termed the co-ordinates of the point.

The features of graphs are indicated in Figure 8.

The use of graphs is illustrated in this chapter and also in the later stages of the book.

The undermentioned points arise when constructing a graph:—

(1) The x and y scales are normally constructed to read from bottom to top and from left to right respectively.

(2) The x and y scales and the lines themselves should be clearly defined, especially when negative values are plotted. Similar considerations apply to a 100 per cent. line, used when percentages are plotted.

(3) The scales, not necessarily equal, should be chosen to adequately show movements in the data.

(4) The scales should be indicated in round figures.

(5) Should it be impracticable to show the zero point on the vertical scale, a 'break' should be indicated. (See Figure 11.)

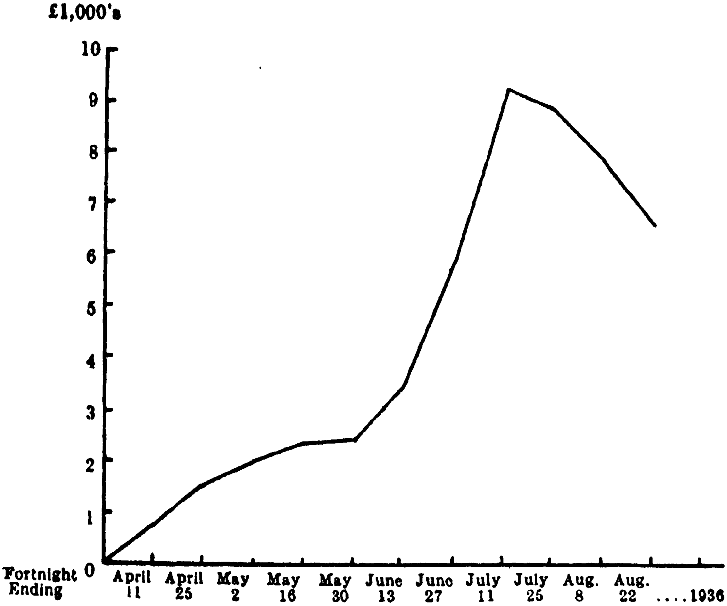
(6) Where two or more series of data are plotted the curves should be distinguished by different colours or types of ruling, and a key to these variations given.

(7) The actual draughtsmanship of the graph will vary according to the information to be plotted. In this connection the graph should be as large as possible consistent with ease of reading and handling and the scale should be such that movements in the data are readily apparent. The latter may be achieved by selecting a suitable vertical scale.

(8) The points representing a series of observations should be indicated clearly, whenever appropriate.

(9) An adequate heading should be supplied.

Figure 9.



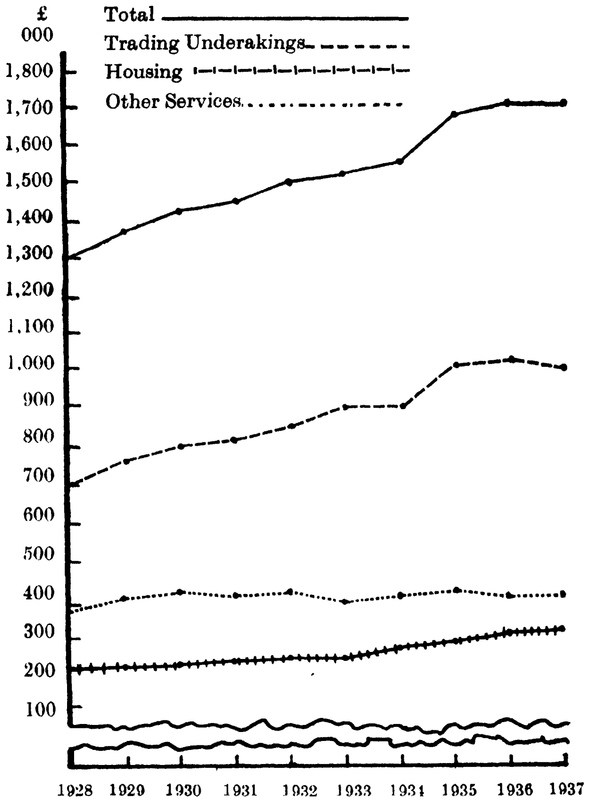
GRAPH SHOWING FORTNIGHTLY COLLECTIONS OF RATES

(*a*) *Example.* Figure 9. Graph showing the actual amounts of rates collected at fortnightly intervals.

DATA

Fortnight ending		Amounts collected £
April	11	750
,,	25	1,500
May	2	1,980
,,	16	2,400
,,	30	2,500
June	13	3,500
,,	27	6,000
July	11	9,400
,,	25	9,000
August	8	8,000
,,	22	6,750
&c., &c.		—
		51,780

Figure 10. GRAPH SHOWING OUTSTANDING LOANS



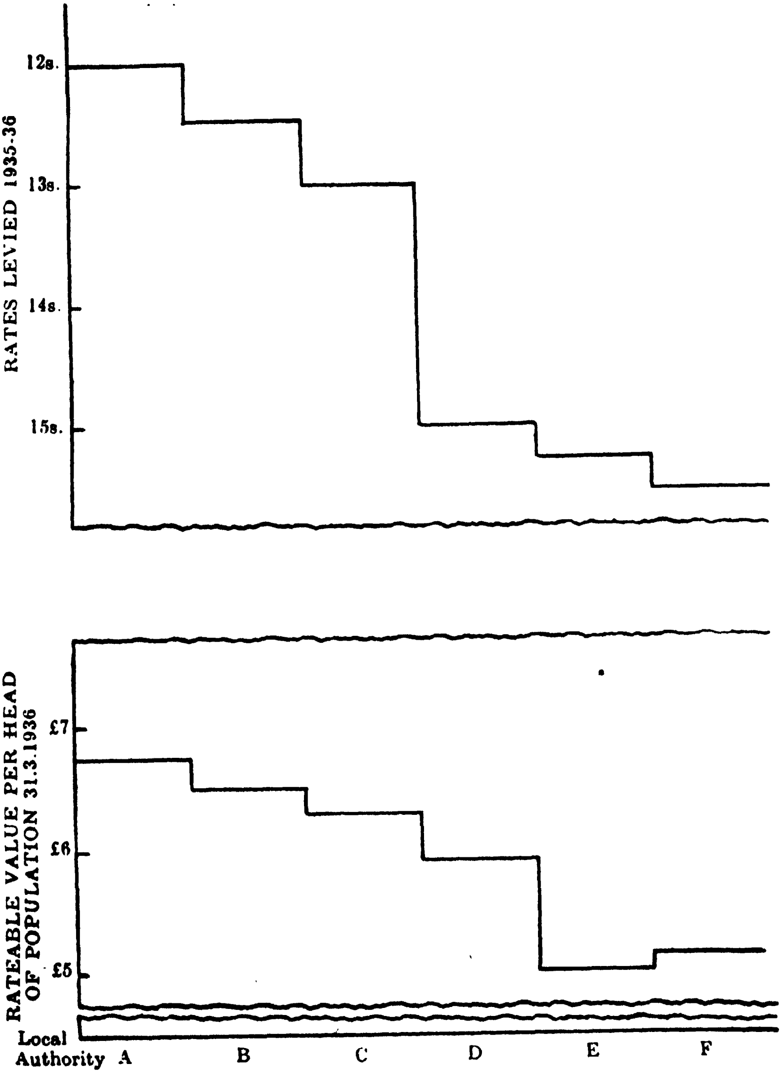
(*b*) *Example.* Figure 10. Graph showing several factors plotted on the same chart and designed to indicate the absolute movements in the data, and its analysis under three headings.

DATA. Loans outstanding at 31st March.

Year		Total	Trading Undertakings	Housing	Other Services
—		000	000	000	000
1927–28	..	1,300	700	210	390
1928–29	..	1,370	760	210	400
1929–30	..	1,435	800	225	410
1930–31	..	1,458	820	238	400
1931–32	..	1,500	850	240	410
1932–33	..	1,530	900	240	390
1933–34	..	1,560	900	260	400
1934–35	..	1,690	1,000	280	410
1935–36	..	1,710	1,010	300	400
1936–37	..	1,705	1,000	305	400

Figure 11.

GRAPH SHOWING COMPARISONS



(*c*) *Example*. Figure 11. Graph, with two vertical scales, reading in different directions, showing comparisons.

DATA

Local Authority	Rateable Value per Head of Population £	Rates Levied s.	d.
A	6.75	12	0
B	6.5	12	6
C	6.3	13	0
D	5.9	15	0
E	5.0	15	3
F	5.2	15	6

(*d*) Where measures of x and y are expressed in different units comparison of the movements of the data may be facilitated by equating the factors to be plotted, either on the basis of y on x (for the first period of the comparison) or by calculating index numbers for both sets of data (Chapter VIII).

In the example given it is desired to compare a series of figures reduced to index numbers with other factors expressed in shillings. The method is to equate the value of y to that of x for period No. 1., viz., 90. Subsequent values of y are similarly equated on the basis of the first period.

Period	Original Value	Equated Value	
2	60	90	$60 \times \frac{90}{60}$
3	58	87	$58 \times \frac{90}{60}$
4	54	81	$54 \times \frac{90}{60}$
	&c.	&c.	

DATA

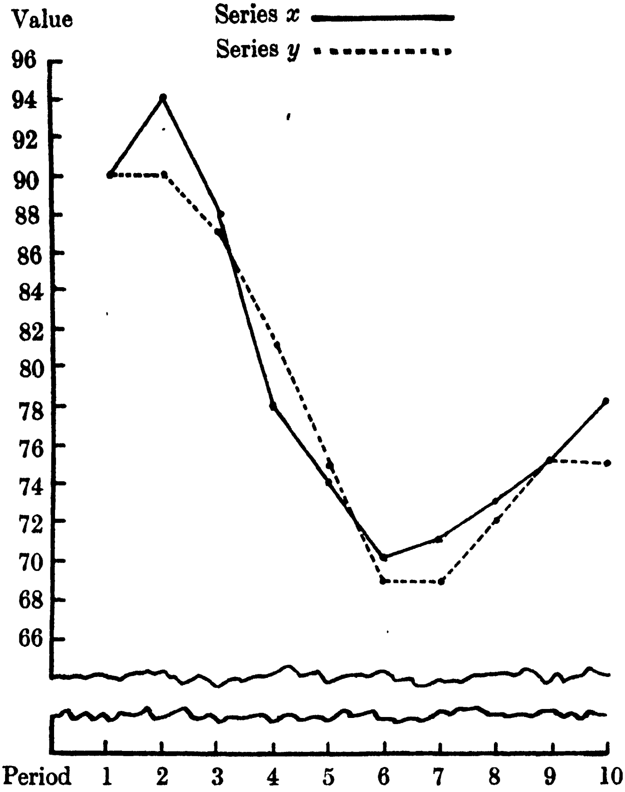
Period	Index (x)	Weekly Wages Actual (y)	Weekly Wages Equated
1	90	60/-	90
2	94	60/-	90
3	88	58/-	87
4	78	54/-	81
5	74	50/-	75
6	70	46/-	69
7	71	46/-	69
8	73	48/-	72
9	75	50/-	75
10	78	50/-	75

The original values of x and the equated values of y are compared in Figure 12.

Figure 12.

GRAPH SHOWING COMPARISONS

(Variables equated at commencing date.)



7. *Logarithmic Curve*, is a curve where, by reducing natural figures to logarithms and plotting these values, the ratios of variation are shown. The use of logarithmic curves is restricted to investigations where the ratio of increase or decrease—the relative movements, and not the actual variation—the absolute movements, is required, e.g. when comparing the increase in loan debt with the increase in population.

The effect of the comparison of ratios of increase and decrease is given in Figure 13, the absolute differences being the same.

DATA

Period	Series (1)	Series (2)
A	1	5
B	4	8
C	2	6
D	3	7
E	2	6
F	5	9

In this example the natural numbers are plotted on semi-logarithmic ruled paper, thus avoiding the conversion of the natural numbers into logarithms.

As a logarithmic curve shows ratios of variations on any part of the scale, it is immaterial on what section of the diagram the facts are plotted. Thus widely differing items, expressed as natural numbers, may be converted to logarithms and plotted in proximity to one another, thus facilitating comparison.

For similar reasons no zero line is given, owing to the very real possibility of inducing fallacious conclusions.

It adds to the value of a logarithmic curve if—

(1) The natural numbers are given, where possible, as well as their values expressed as logarithms.

(2) Bars are given indicating the ratio of change.

Another example of a logarithmic graph is given in Figure 14.

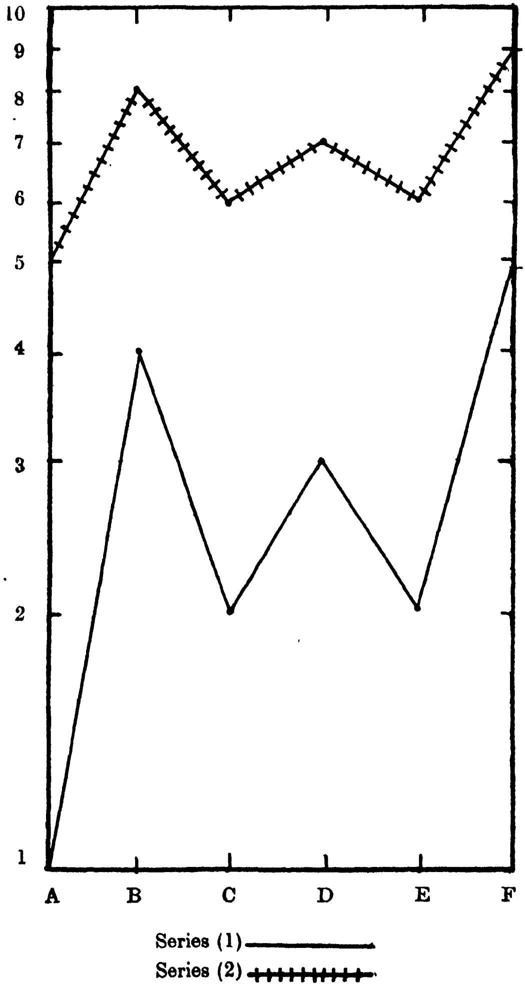
DATA

Year	Net Loan Debt £1,000s	Logs
1928	50	4.6990
29	53	4.7243
30	60	4.7782
31	55	4.7404
32	50	4.6990
33	54	4.7324
34	59	4.7709
35	75	4.8751
1936	88	4.9454

Figure 13.

LOGARITHMIC GRAPH

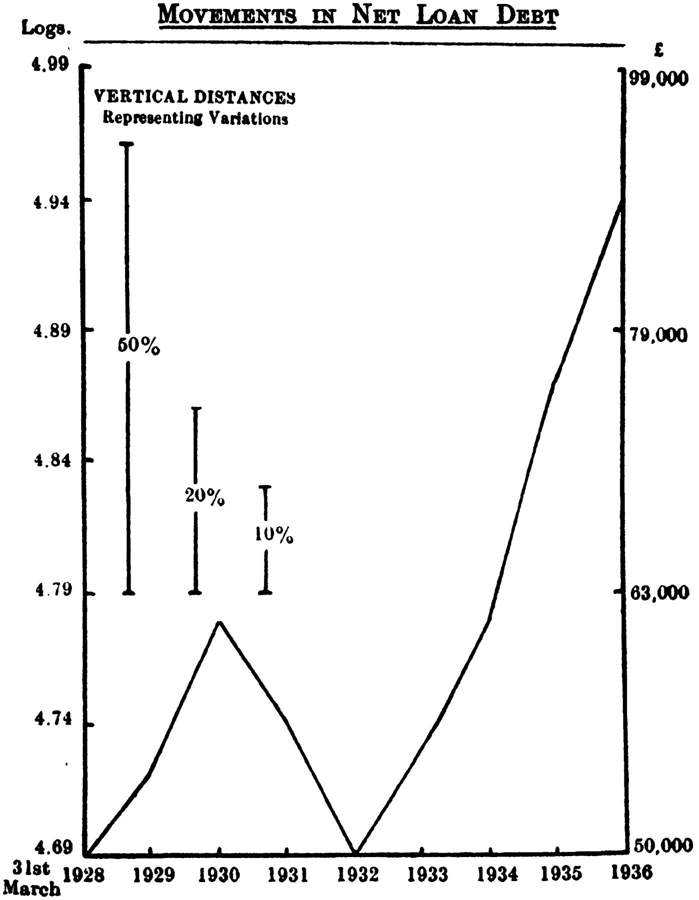
Graph illustrating the use of the logarithmic scale in showing ratios, the absolute movements in the two series being the same.



Note.—1. Figure 13 is more correctly a semi-logarithmic graph, where one variable (usually y) is plotted on the logarithmic scale and the other (x) on the natural scale.

Figure 14.

LOGARITHMIC GRAPH SHOWING MOVEMENTS IN NET LOAN DEBT



Note.—2. Figure 14. Owing to the difference in base of the two sets of numbers, both scales of y will not progress in a uniform manner.

3. Figure 14. The ascertainment of the length of the bar is as follows:—
   (*a*) An increase of 50 per cent. is effected by multiplying the number by 1.5.
   (*b*) The log of 1.5 is log .17.
   (*c*) The distance representing log .17 on the vertical scale, will indicate an increase of 50 per cent.

Similarly with any increase it is desired to illustrate. Other points arising in the draughtsmanship of these charts are given below.

(1) The use of specially ruled paper greatly facilitates the preparation of logarithmic graphs.

(2) As the addition of .301 to the logarithm of any number gives the logarithm of twice that number, where large fluctuations arise it may assist in reading the graph if this factor is incorporated in the scale adopted, e.g. where the increase in the logarithm .301 is represented by say 1 inch in the scale, a variation of 1 inch signifies a variation of 100 per cent. in the natural numbers of the quantity or value represented.

(3) The values of y are obtained by
   (*a*) plotting the natural values of y on semi-logarithmic ruled paper. Example, Figure 13;
   (*b*) plotting the logarithms of y. Example, Figure 14.

8. *Frequency Curves.*

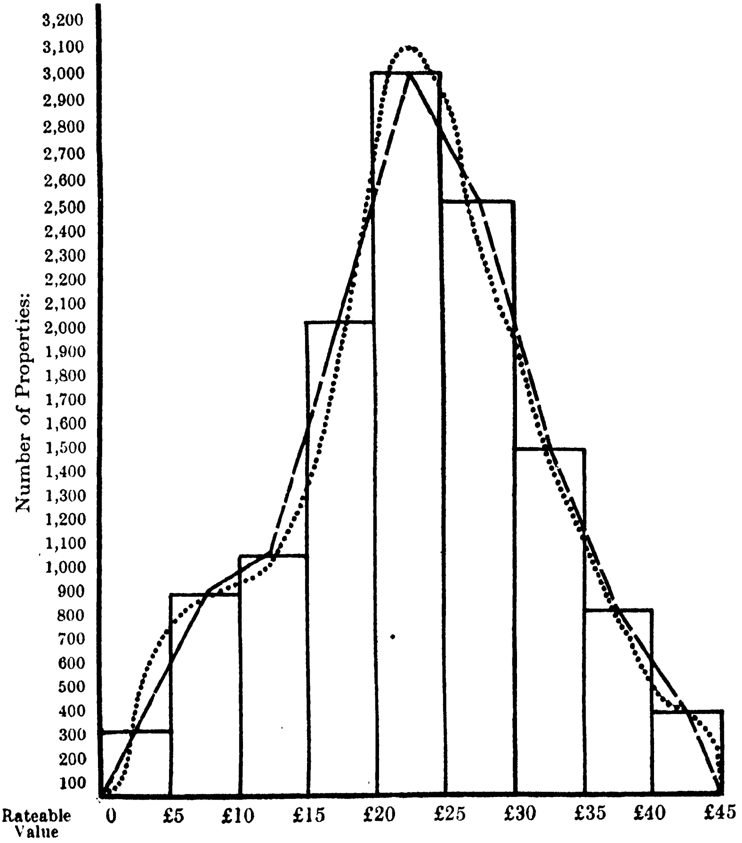
A frequency curve is the diagram formed by plotting a frequency table in the form of a graph. It is used for distributive comparisons, i.e. for investigating the relation of items in a frequency table to each other and to other tables plotted on the same chart. Although information may be deduced from, or by, comparing frequency tables, the investigation is materially assisted if the data is pictorially represented. Further, an average (Chapter VI) may be useful in certain methods of inquiry, e.g. in comparing groups; but an average may not be any particular item and in any case, irregularities (which it may be desired to measure) tend to be negatived. Again, it may be necessary to compare the relation of items to one another or to the average. In these cases a frequency curve supplies the best medium of investigation.

In preparing a frequency curve it is usual for class intervals or the magnitude of the items to be shown on the horizontal scale, and the frequency (i.e. the number in the class interval or the number of items) on the vertical scale. If the class interval is a whole number the data can be plotted, and the points joined, so as to form a frequency curve, without preliminary preparation (subject to 'smoothing' where necessary, *post*). Should the class interval be otherwise than a whole number, e.g. under 10, over 10 and under 20, the following operations are necessary:—

(1) Plot the data so that the frequency and class intervals form rectangles the areas of which are proportional to the frequencies. (See Example, Figure 15, continuous lines.)

Figure 15.

FREQUENCY CURVE SHOWING THE NUMBER OF PROPERTIES UNDER THE HEADING OF RATEABLE VALUE



(2) Form a 'frequency polygon', i.e. the figure produced by joining the mid-points of the class intervals of a frequency table plotted as a graph. (See Example, Figure 15, broken lines.)

(3) Form a 'smoothed frequency curve', i.e. the figure formed by 'rounding off' a frequency polygon to form the assumed ideal curve. (See Example, Figure 15, dotted lines.)

The data used in Example, Figure 15, is as follows:—

Rateable Properties between £5 and £45 Rateable Value.

					Frequency
Rateable Value not exceeding £5					276
Exceeding	£5	and not	exceeding	£10	843
,,	10	,,	,,	15	1,007
,,	15	,,	,,	20	1,984
,,	20	,,	,,	25	3,010
,,	25	,,	,,	30	2,476
,,	30	,,	,,	35	1,540
,,	35	,,	,,	40	781
,,	40	,,	,,	45	352

A bell-shaped normal frequency curve arises when the data is evenly distributed. Such a curve always results in connection with natural data, but does not necessarily follow when economic data is plotted.

The term 'normal curve of error' is also applied to a normal frequency curve. This term arises from the fact that a group of items has a position of greatest density (the situation of the normal or most common item, i.e. the mode or norm, Chapter VI), and this curve is a method of investigating the variations from the average. 'Error' does not refer to a mistake, but the variation of individual items from the normal or most common item.

See Chapter XI and Figures 37 and 38.

Smoothing is justifiable—

(1) as the class boundaries are dependent, not on the actual data, but on the class intervals selected. If a smaller class interval was utilised the line followed by the frequency polygon would more nearly approach the assumed ideal curve.

(2) In dealing with very large numbers, e.g. the number of children in England analysed under the heading of age:—

(*a*) the plottings would be so numerous as to form a continuous curve;

(*b*) where sampling is adopted the law of statistical regularity would apply, when the abbreviated data should have the characteristics of the whole.

The extent to which smoothing may be adopted may be indicated as follows:—

(*a*) Natural data—completely regularise the curve.

(*b*) Economic data—eliminate minor irregularities only.

Experience has shown that the contracted method of selecting suitable samples and class intervals and smoothing the resulting polygon is amply justified; although its result is to show tendencies and hide irregularities appearing in the data under review.

Note.—(1) The area of the smoothed curve should equal the area of the rectangles.

(2) The extension of the curve at the peak is justifiable as there is every likelihood of a large number of items being situated at the mid point of the class interval.

The term 'histogram' is applied to a curve or diagram showing the frequency distribution of data, i.e. to show the frequency at which items occur in a series, and 'historigram' where the curve shows changes over a period of time, e.g. rateable values over a period of years, &c.

In constructing a histogram the magnitudes are plotted on the horizontal scale, while in the case of the historigram the vertical scale is utilised for this purpose.

Skewness.—A symmetrical bell-shaped curve results from the plotting of natural phenomena, but does not necessarily follow when economic or other such data is plotted. 'Skewness' is the term applied to this absence of symmetry. It is derived from the irregular shape of a frequency curve plotted from data that is not evenly distributed, i.e. the frequencies on either side of the position of greatest ordinate (the 'mode'—Chapter VI), are uneven.

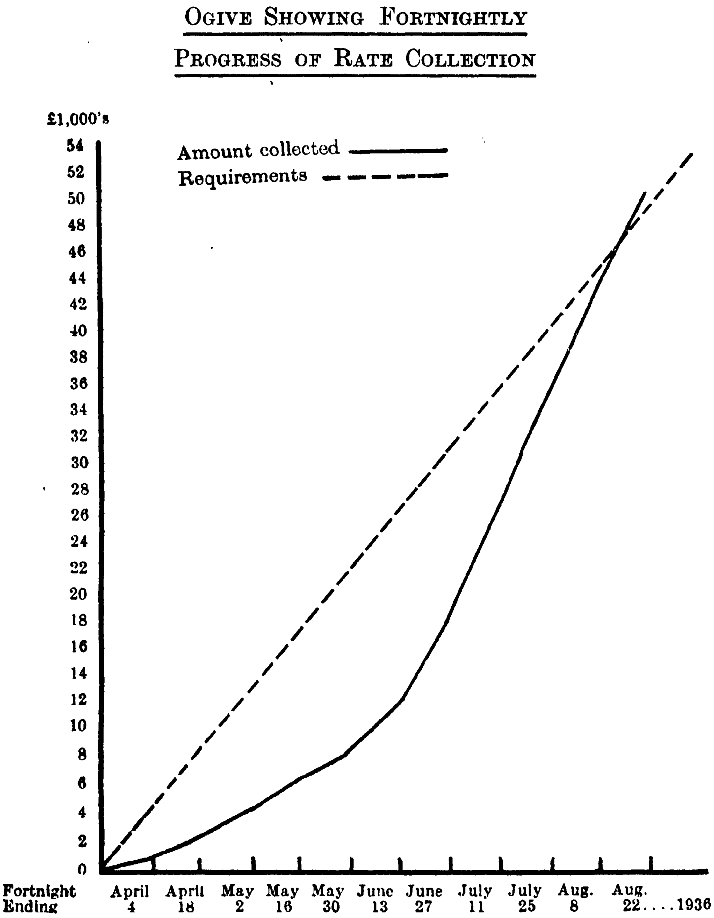
The measurement of skewness is often of importance in statistical investigations, when it is desired to make comparisons in the fluctuations in the distribution of sets of data (Chapter IX). Examples of skewness are given in Figures 29 and 30.

9. *Ogives.*

An ogive, cumulative frequency or summation curve is a curve plotted from a cumulative frequency table and is designed to show progressive comparisons, i.e. the cumulative frequency of data. It is used chiefly in the case of a continuous inquiry, e.g. a chart showing the progress of rate collections, and also for determining the position and size of certain averages (e.g. the median, quartiles and deciles: Chapter VI and Figure 20). The curve should be smoothed as in the case of a frequency curve.

Example. Figure 16.—Ogive showing the fortnightly progress of rate collections.

Figure 16.



DATA

Fortnight ending		Amount Collected £	Cumulative Collections £
April	4	750	750
,,	18	1,500	2,250
May	2	1,980	4,230
,,	16	2,400	6,630
,,	30	2,500	9,130
June	13	3,500	12,630
,,	27	6,000	18,630
July	11	9,400	28,030
,,	25	9,000	37,030
August	8	8,000	45,030
,,	22	6,750	51,780
		51,780	

10. *Lorenz Curve*, which is a variation of the ogive, is a curve designed to show the inequality in the distribution of the particular subject of inquiry, or alternatively its concentration over the group of data. This is achieved by—

(1) Constructing a cumulative frequency table.

(2) Expressing the factors as percentages.

(3) Plotting the figures so obtained on a graph upon which the 100 per cent. division on the horizontal scale coincides with the 0 per cent. division on the vertical scale. A line joining the 100 per cent. point on the vertical scale with the 0 per cent. point on the horizontal scale will form the line of equal distribution. The curve thus illustrates the deviations of sets of data—

(*a*) from regularity (the line of equal distribution);

(*b*) as compared with each other.

A disadvantage of this form of curve is that the curves tend to coincide at the extremes. This objection can be met by utilising a much larger scale and plotting the extremes on separate charts.

Example. Figure 17.—Lorenz curve showing the distribution of properties on the basis of rateable value. (Curves showing comparison with other local areas could also be plotted on the same chart.)

DATA

Cumulative Frequency and Cumulative Percentage Table

Rateable Value	Average Rateable Value £	Percentage	Frequency	Cumulative Frequency	Percentage
Not exceeding £10 ..	5	6	961	961	8
Exceeding—					
£10 & not exceeding £20	15	17	2,460	3,421	30
20 " " 30	25	29	5,712	9,133	80
30 " " 40	35	41	1,503	10,636	93
40 " " 50	45	53	546	11,182	98
50 " " 60	55	64	95	11,277	99
60 " " 70	65	76	51	11,328	99
70 " " 80	75	89	22	11,350	99
80 " " 90	85	100	5	11,355	100
			11,355		

Figure 17.

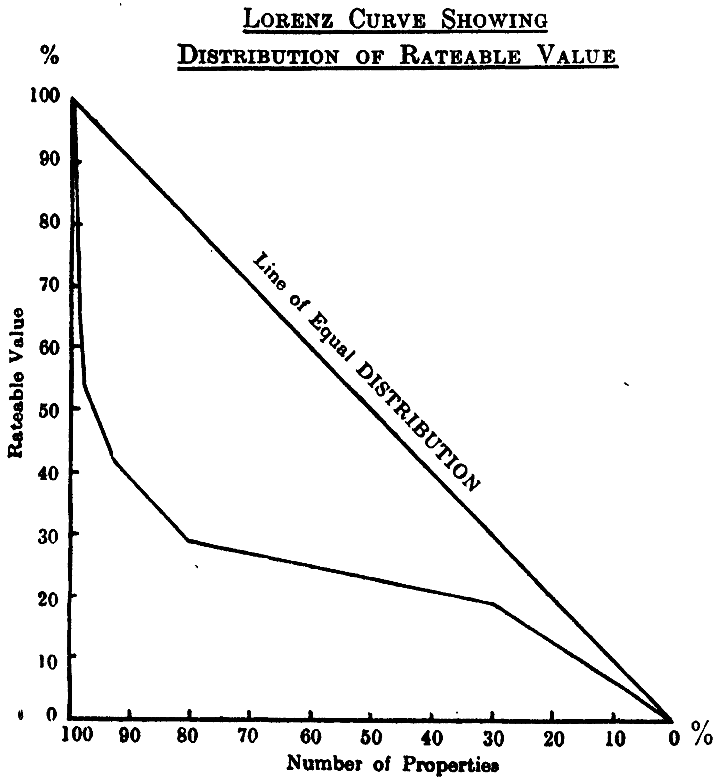


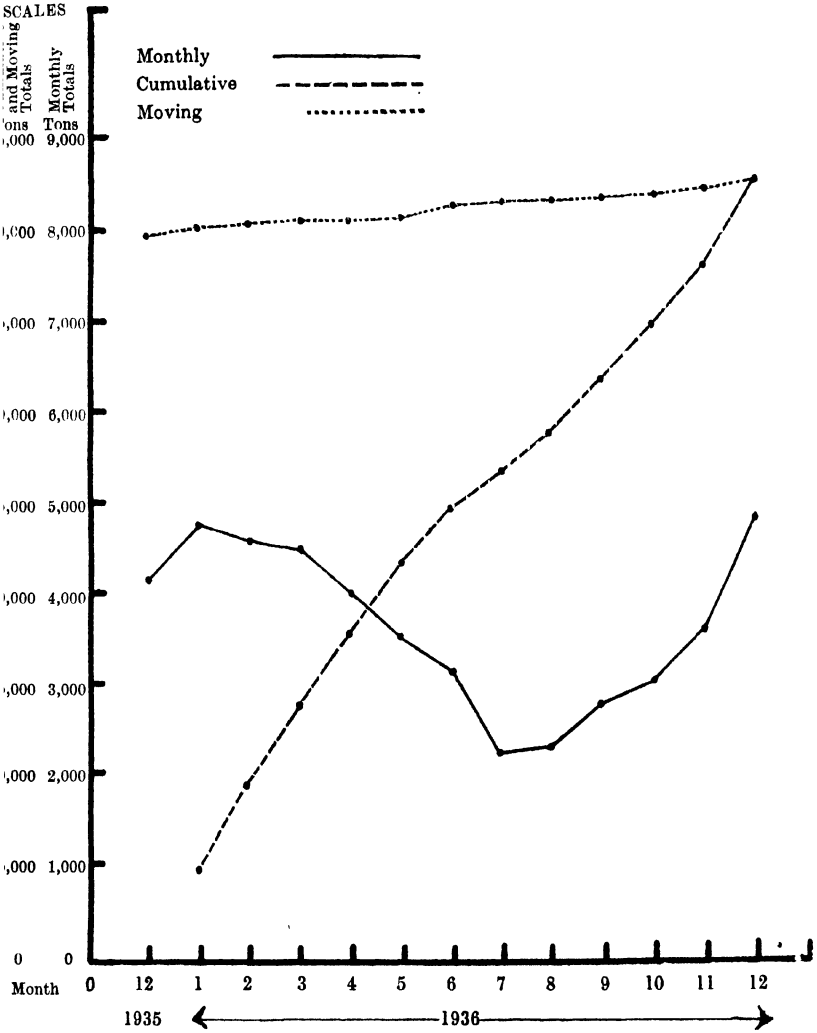
Figure 18.

Z-CHART

GAS UNDERTAKING—SALE OF COKE

1936

TOTALS	Moving	39,996	40,030	40,129	40,373	40,401	40,568	41,168	41,368	41,371	41,573	41,791	41,991	42,540
	Cumulative	39,996	4,700	9,215	13,625	17,619	21,119	24,219	26,422	28,672	31,386	34,292	37,822	42,540
	Monthly	4,169	4,700	4,515	4,410	3,994	3,500	3,100	2,203	2,250	2,714	2,906	3,530	4,718



11. *Z-Charts.*

The Z-chart is an historical record composed of three curves showing:—

1. The data relating to each period.
2. The cumulative figures.
3. A moving total.

The periods of time are set out on the x axis. The y axis has two scales, one for the data relating to each period and the other (the smaller) for the cumulative figures and the moving total.

In the example given, Figure 18 and supporting table, which relates to the sales of a residual (coke) of a gas undertaking, the period selected is one month, and the chart covers a year's working. The cumulative figures give the total sales to date, commencing on January 1st. The moving totals give the total figures for a year to date, but commencing 12 months prior, e.g. the figure for January, 1936, in example below is the total for the year 1935, minus the figures for January, 1935, plus January, 1936 (39,996 − 4,666 + 4,700 = 40,030).

The cumulative and moving total figures meet at the end of the total period covered, in this case one year.

It is convenient to adopt the undermentioned proportions in regard to the relation of the scales:—

Monthly figures with yearly moving totals— 5 : 1
Weekly figures with yearly moving totals—20 : 1
Daily figures with monthly moving totals—10 : 1

Example. Data relating to a gas undertaking—Sales of coke, in support of Figure 18.

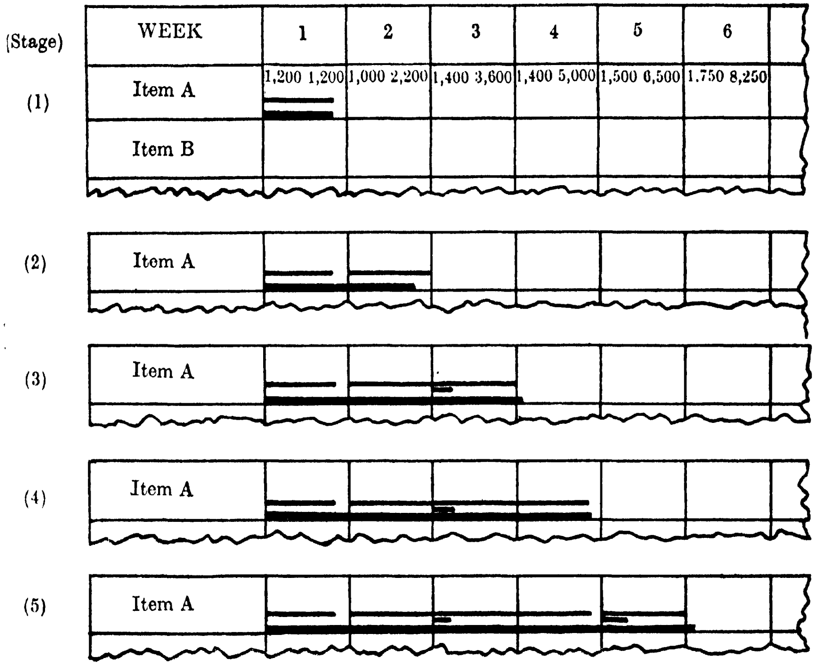
	1935			1936		
Month	Monthly (tons)	Cumulative (tons)	12 months moving total (tons)	Monthly (tons)	Cumulative (tons)	12 months moving total (tons)
January	4,666	4,666	40,666	*4,700*	*4,700*	*40,030*
February	4,416	9,082	40,582	*4,515*	*9,215*	*40,129*
March ..	4,166	13,248	40,748	*4,410*	*13,625*	*40,373*
April ..	3,966	17,214	40,714	*3,994*	*17,619*	*40,401*
May ..	3.333	20,547	40,547	*3,500*	*21,119*	*40,568*
June ..	2,500	23,047	40,547	*3,100*	*24,219*	*41,168*
July ..	2,003	25,050	40,550	*2,203*	*26,422*	*41,368*
August ..	2,247	27,297	40,297	*2,250*	*28,672*	*41,371*
September	2,512	29,809	40,309	*2,714*	*31,386*	*41,573*
October ..	2,688	32,497	39,997	*2,906*	*34,292*	*41,791*
November	3,330	35,827	39,827	*3,530*	*37,822*	*41,991*
December	4,169	39,996	39,996	*4,718*	*42,540*	*42,540*
	39,996			*42,540*		

The chart covers the results (monthly) for a period of one year, plus the last month of the previous year. The inclusion of the latter in this type of chart is valuable in affording continuity and in indicating the trend of the monthly and annual moving totals at the commencement of the new yearly period.

Figure 19.

GANTT PROGRESS CHART

Electricity or Gas Undertaking—Meter Reading



(The above represents successive stages in the compilation of the same chart.)

12. *Gantt Charts.*

The Gantt chart is an historical record, composed of two lines showing:—

1. The data relating to each period—indicated by a thin line.
2. The cumulative figures—indicated by a thick line.

Both sets of figures are expressed as percentages of a quota or predetermined standard.

The chart is divided into equal sections representing a period of time and these periods are set out horizontally across the chart. The chart admits of the inclusion of several items, e.g. sales and/or production of various items, stages in constructional programmes, or any process capable of analysis over a period of time. The items are set out vertically on the left of the chart.

The standard or quota for each period is indicated in the top left-hand corner of each section and the cumulative quota figures on the right.

The actual results for each period and cumulatively are indicated by the length of two horizontal lines drawn across the appropriate section, when the width of each section represents 100% performance during that period.

In the example given, Figure 19, and supporting table, which relates to the progress made in the reading of meters in an electricity or gas undertaking, the period selected is one week, and assuming continuous billing in conjunction with the rendering of quarterly accounts, the chart would cover a total period of 13 (or allowing one week for contingencies), perhaps 12 weeks.

For the sake of clarity one item only (Item A—Meter Reading) is illustrated in example below, and the chart (Figure 19).

Example. Data relating to an electricity or gas undertaking—Meter Reading, in support of Figure 19.

Weekly Results. (Thin Line).

Week	Quota	Actual	% of Quota	Plotting
1	1,200	1,000	83	Line extending to 83% of width of section.
2	1,000	1,005	100	„ „ 100% „ „
3	1,400	1,680	120	„ „ 100% + 20% „ „
4	1,400	1,260	90	„ „ 90% „ „
5	1,500	1,950	130	„ „ 100% + 30% „ „
6	1,750			
7 &c.				

Cumulative Results. (Thick Line.)

Week	Quota (Cumulative)	Actual to Date	Per cent.	Plotting.
1	1,200	1,000	83%	Line extending to 83% of width of section.
2	2,200	2,005	* 100% of Week No. 1, plus 80% of Week No. 2	Line extending to 100% Week No. 1, plus 80% of Week No. 2.
3	3,600	3,685	100% of Weeks Nos. 1 to 3, plus 6% of Week No. 4	Line extending to 100%, Weeks No. 1 to 3, plus 6% of Week No. 4.

Week	Quota (Cumulative)	Actual to Date	Per cent.	Plotting.
4	5,000	4,945	100% of Weeks Nos. 1 to 3, plus 96% of Week No. 4.	Line extending to 100%, Weeks Nos. 1 to 3, plus 96% of Week No. 4.
5	6,500	6,895	100% of Weeks Nos. 1 to 5, plus 5% of Week No. 6	Line extending to 100% Weeks Nos. 1 to 5, plus 5% of Week No. 6.
6	8,250			
7 &c.				

* i.e. 1,200 to cover Week No. 1 = 100%.
805 on account of 1,000 (Week No. 2) = 80%.

Note.—Figure No. 19 illustrates the successive stages of the chart. In practice, charts are not re-drafted for each period, but the new information arising is added from period to period, by extending the horizontal lines in the proportion the actual figures bear to the quota.

Chapter VI

AVERAGES

Averages — Characteristics—Advantages—Disadvantages : Arithmetic Average—Advantages—Disadvantages : Weighted Arithmetic Average: Moving Average: Descriptive Average: Progressive Average: Geometric Mean—Advantages—Disadvantages: Harmonic Mean—Advantages—Disadvantages: Percentages: Ratios: The Mode—Advantages—Disadvantages: The Median—Advantages—Disadvantages: Quartiles: Deciles: Percentiles: Further Examples.

THE scientific selection of a type or the calculation of an average, i.e. a typical example, summarising the whole of the data, is often resorted to, chiefly for making comparisons. For example, a statement of the rateable values of a number of local areas per head of population, is likely to be more useful and intelligible than a mere statement of their total rateable values.

The following points should be considered in calculating or selecting an average.

(1) It should be clearly defined and not admit of misconstruction.
(2) It should be characteristic of the whole data under consideration and, if possible, an actual example.
(3) It should be chosen from a large range of data so as to nullify the effect of abnormalities and to avoid undue fluctuations where further matter is introduced.
(4) Preferably it should be calculated on a simple basis.
(5) The method of selection or calculation should be apparent or else disclosed by a footnote or in supporting memoranda.
(6) It should admit of mathematical treatment.

No known average possesses the whole of the foregoing properties, so in practice the most suitable type is selected having regard to the data available, and the mathematical use to which it is desired to subject the average.

The relàtive advantages and disadvantages of the use of averages are given below:—

Advantages.

(1) An average forms a convenient medium for representing a particular aspect or aspects of a large group or mass of data.
(2) An average taken from sample data is, in the majority of cases, as accurate as an average calculated from the complete data.
(3) The influence of extremes or abnormalities is minimised.
(4) An average forms a more convenient method of comparison than a complete group or table.

(5) An average is a useful basis for comparing past and present items and for forming the basis for future comparisons.

Disadvantages.

(1) An average is chacteristic of the data as a whole and does not show the trend. For example, the average population of an area for the past four years may be 19,015, when the facts are as follows:—

Example 9.

(*a*) Year W	20,061	or (*b*) Year W	16,864
X	19,976	X	18,047
Y	18,001	Y	19,103
Z	18,022	Z	22,046
	4)76,060		4)76,060
	19,015		19,015

(2) An average may not be an actual item.

(3) An average not supported by actual figures may lead to fallacious conclusions.

(4) Arithmetical inaccuracies are often made when calculating averages.

(5) The incorrect use of averages is prevalent, i.e. the wrong type of average, although mathematically correct, may be used.

1. *The Arithmetic Average* (Arithmetic Mean).

The arithmetic average is, perhaps, the simplest and most generally used of the various averages. It is calculated by dividing the sum of a series of items under review by the actual number of items.

Formula:

Let a = the arithmetic average.
m = the items in the series.
n = the number of items in the series.

$$\text{Then } a = \frac{\Sigma m}{n}$$

Note.—Σ = the sum of

Example 10.

Items
651
746
863
740
915
5)3,915
Arithmetic Average = 783

Alternatively as the algebraic sum of the deviations on either side of the arithmetic mean is zero, the mean may be ascertained by—

(1) Assuming an approximate average.
(2) Set out the difference (+ and −) of each item from the assumed average.
(3) Total and divide the differences by the number of items.
(4) Adjust the assumed average by adding or subtracting the sum ascertained in item (3).

when the result will give the arithmetic average.

Example 11.

Using the same data as above, assume the average to be 800. Then the arithmetic average will be:—

$$800+\left\{\begin{array}{rr} + & - \\ & 149 \\ & 54 \\ 63 & \\ & 60 \\ 115 & \\ \hline +178 & -263 \end{array}\right\}\div 5$$

$$=800+\left\{\frac{178-263}{5}\right\}=800-17$$

$$=783$$

It has these *advantages*:—

(1) Its calculation is easy and generally understood.
(2) The data needs very little preparation, e.g. it need not be arrayed.
(3) It may often be calculated from data incomplete for other purposes, e.g. if the magnitude of the items is unknown, but their aggregate and number known, the average may be calculated.
(4) Each item is given due weight.
(5) It may be used as the basis for further calculations.

And the following *disadvantages*:—

(1) It may be considerably affected by extremes.
(2) It may not be an actual item.
(3) It cannot be ascertained by inspection.
(4) Numerical measurement is a necessary preliminary to its calculation.
(5) Unless supported by the figures from which it was calculated, its use may result in fallacious deductions.

2. *Weighted Arithmetic Average.*

The term 'weighted' is applied when an average is calculated on data arranged in groups and the frequency, i.e. the number of items in each group of the data, taken into account. In this case the class interval is multiplied by the frequency, and the results so obtained are totalled and divided by the sum of the frequencies, i.e. the weights.

A more correct result is obtained by using weights than by calculating a simple arithmetic average, although in the majority of cases where the range of items is small and the frequencies large, there is very little difference between a simple and a weighted average.

It will be apparent, however, that where weights are used, the greatest accuracy possible must be obtained in the values or magnitudes of the items. The weights, however, may be approximated.

Formula.

Let a, b, c = items in a series,

w_1, w_2, w_3 = the weights applicable to the items,

then the weighted arithmetic average

$$= \frac{(a \times w_1) + (b \times w_2) + (c \times w_3)}{w_1 + w_2 + w_3} = \frac{\Sigma m_w}{\Sigma w}$$

Alternatively, where f represents the frequency of the individual items:—

$$\frac{\Sigma\, f\, m}{\Sigma f}$$

Example 12.

It is required to find the average rate of interest on loans outstanding, i.e. the weighted arithmetic average of the various rates of interest.

Average Rate of Interest on Loans

Rate of Interest	Loan Debt	Product
%	£	£
4	100,610	402,440
3½	54,026	189,091
3	86,167	258,501
3¾	24,171	78,556
	264,974	928,588

Then the average rate of interest is

$$\frac{928{,}588}{264{,}974} = \text{£}3 \ \ 10\text{s. } 1\text{d. per cent.}$$

3. *The Moving Average*, is an arithmetic average of data arising over a period and is calculated by replacing the first item in the average by the newly arising item. It gives the trend of the subject under inquiry and tends to nullify the effect of fluctuations and short term influences. It is especially useful in considering such matters as the trend of prices, sales, &c., although care is needed in deciding the

period of time to be used as the basis of the average, the chief factor arising being the nature of short term fluctuations and their influence on the data under review.

Example 13.

5-YEAR MOVING AVERAGE

Year	Average Price		Moving Average	
	s	d	s	d
1926	5	10	—	
1927	5	8	—	
1928	5	7	—	
1929	5	7	—	
1930	5	6	5	7⅗
1931	5	6	5	6⅘
1932	5	5	5	6⅕
1933	5	2	5	5⅕
1934	5	1	5	3⅗
1935	4	9	5	2⅕

4. *Descriptive Arithmetic Average.*

The term ‘descriptive’ is applied where an arithmetic average is calculated on data which is incomplete as to the actual number of items of different magnitudes. The average cannot be regarded as a reliable basis for further calculations, unless there is a reasonable possibility of the data being equally distributed between the class limits.

Example 14.

The descriptive average of 850 hereditaments between £10 and £20 rateable value would be £15, i.e. the most likely average.

5. *Progressive Average.*

Where it is not practicable to use a moving average, e.g. where the period is not long enough, a progressive average may be calculated. In this case an average is calculated from the information available to date, new factors being introduced from time to time.

Example 15.

PROGRESSIVE AVERAGE

Year	Average	Progressive Average	
1930	£1,000	—	
1931	1,200	1930–31	£1,100
1932	1,600	1930–32	1,267
1933	1,400	1930–33	1,300
1934	1,700	1930–34	1,380
1935	1,500	1930–35	1,400

6. *Geometric Mean.*

This average is calculated by multiplying the items in the series together and extracting the root, the basis of which is the number of items, i.e. the nth root of the product of the n items in the group. Or, where logarithms are used, the logarithm of the geometric mean of a series is the arithmetic mean of the sum of the logarithms of the items in the series.

Formula.

Let a, b, and c . . . be items in a series, and n the number of items.

Then the geometric mean $= \sqrt[n]{abc} \ldots$

Example 16.

It is required to find the geometric mean of the items 4, 10 and 12.8.

Applying the formula, the geometric mean is

$$\sqrt[3]{4 \times 10 \times 12.8} = 8.$$

The principle of weighting may also be applied to this average, when the formula will be modified as follows:—

Let a, b and c . . . be items in a series and the weights applicable, w_1, w_2, w_3 . . .

Then the weighted geometric mean=

$$\sqrt[\Sigma(w)]{a^{w1} \times b^{w2} \times c^{w3}} \; \ldots .$$

The items are raised to a power equal to the required weights and the root to be extracted is equivalent to the sum of the weights.

Advantages.

(1) It measures the ratio of change or variation in the data.

(2) It gives little weight to extreme items and is thus smaller than the arithmetic average.

(3) It is especially valuable compared with the arithmetic average, when the latter is not representative owing to the presence of a few items of considerable magnitude.

Disadvantages.

(1) Its calculation is involved, usually requiring the use of logarithms.

(2) It is not commonly understood.

7. *Harmonic Mean.*

The Harmonic Mean of a series is the reciprocal of the arithmetic mean of the reciprocals of the items and is calculated by—

(1) Ascertaining the reciprocals of the items.

(2) Ascertaining the arithmetic mean of the total of the reciprocals obtained in operation (1).

(3) The Harmonic Mean is the reciprocal of operation (2).

Formula.

Let a, b and c . . . be items in a series.
n = the number of items.
h = the harmonic mean.

$$\text{Then } \frac{n}{h} = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \ldots \text{ or } \frac{n}{\Sigma\left(\frac{1}{m}\right)}$$

When weights are used the formula becomes:—

$$\frac{\Sigma w}{\Sigma \frac{w}{m}}$$

Example 17.

It is required to find the harmonic mean of the items 4, 10 and 12.8. Then the harmonic mean is calculated by:—

$$\frac{3}{h} = \frac{1}{4} + \frac{1}{10} + \frac{1}{12.8} = 7.007$$

Advantages.

(1) It measures the ratio of change or variation in the data.

(2) It gives very little weight to extreme items, and is smaller than geometric mean.

Disadvantages.

(1) Its calculation is involved.

(2) It is not commonly understood.

8. *Percentages.*

A percentage is the term applied where the relationship of the figures to be compared is expressed on the base of 100, e.g. the relationship of 143 to 130 is 110 per cent.

$$\left(\text{i.e. } \frac{143}{130} \times 100\right)$$

Advantages.

(1) It is simple to calculate.

(2) It is generally understood.

Disadvantages.

(1) Fallacious results may be obtained when percentages are used, the bases of which are not comparable.

(2) Percentages of numbers less than 100, i.e. when items are raised to the base of 100, although arithmetically correct, may give false impressions, for example, the deaths of contacts with cases of smallpox may be given as 50 per cent. in two local areas, when the actual figures are 1 out of 2 and 50 out of 100. In a comparison involving the figures given, ratios or the actual figures would be preferable to the use of percentages.

(3) Incorrect results may be obtained by averaging percentages. In such cases the correct percentage is calculated on the sum of the figures under review.

Example 18.

	Rates Collected (1) £	Voids and Irre-coverables (2) £	Percent-ages of (2) to (1) (3)
1932–3	50,000	2,500	5
1933–4	61,400	3,377	5.5
1934–5	55,000	2,337	4.25
1935–6	57,500	2,760	4.8
Total	£223,900	£10,974	19.35
Average ..	£55,975	£2,743	4.837

The percentage should be calculated on the actual totals or the averages and not by dividing 19.35 by 4, when the 'total' and 'average' lines would read as follows:—

Total	£223,900	£10,974	4.901%
Average ..	£55,975	£2,743	

(4) Ambiguity in the use of percentages is prevalent, i.e. the base is not apparent, e.g. wages rose by 10 per cent., then fell by 8 per cent. Does the 8 per cent. reduction apply to the basic figure or the basic figure + 10 per cent.?

The base per mille is also commonly employed, especially in the case of vital statistics. For example, the birth rate of an area is calculated as follows:—

Example 19.

Live births during the year = 1,552

Estimated mid-year population = 101,435

Then the birth rate $= \frac{1,552}{101,435} \times 1,000 = 15.3.$

9. *Ratios.*

A ratio is a direct comparison of the facts under consideration, and unlike percentages, is not founded on an artificial base, e.g. 143 : 130, 1 : .909.

The advantages and disadvantages of ratios are similar to those of percentages, subject to the following:—

(1) Ratios are more akin to the actual figures and fallacious deductions may not be so readily obtained.

(2) The possibility of ambiguity is not so great.

10. *The Mode* (or Norm).

The mode is the predominant item in a group and is situated at the position of greatest numerical density.

The mode may be ascertained—

(1) *By Inspection* from a frequency table or curve, it being situated on the x axis at the position of greatest ordinate, i.e. the peak of the curve.

Example. See Figure 15 and its supporting table.

The mode is situated between £20 and £25.

(2) *By Grouping.*—Items are grouped and re-grouped until a stage is reached when the point of the greatest frequency is not altered by adjustments of the grouping. The mode is situated at the smaller group common to each of the larger groups.

Example 20.

Value	Number			
1	400	1,355		2,310
2	955		1,910	
3	955	1,420		
4	465		700	1,075
5	235	610		
6	375			

The largest frequency 1,420 1st grouping contains values 3, 4

„ „ „ 1,910 2nd „ „ „ 2, 3

„ „ „ 2,310 3rd „ „ „ 1, 2, 3

Value 3 is the mode as it is the smaller group common to each of the larger groups.

(3) The position of the mode in a group can be ascertained approximately by means of the undermentioned formula:—

Let Z = the mode.

L = the lower limit of the class in which the mode is situated (i.e. the model class).

c = the class interval.
f_1 = the frequency of the items in the adjacent lower class.
f_2 = the frequency of the items in the adjacent higher class.

$$\text{Then } Z = L + \frac{f_2 c}{f_2 + f_1}.$$

Example 21.

Adopting the figures from the frequency table on page 72 and applying the formula—

$$Z = 20 + \frac{2{,}476 \times 5}{2{,}476 + 1{,}984} = £22 \quad 15s. \ 6d.$$

Note.—

1. The table shows rateable values. These being calculated to the nearest £1 the mode would be approximated to £13.
2. It is possible for two or more modes to arise in practice, where individual groups of equal frequency occur.

The advantages and disadvantages of the mode as an average are given below:—

Advantages.

(1) In many cases it can be obtained by inspection.
(2) It is an actual item, in fact, the predominant one.
(3) It is not influenced by the presence of a small number of extreme items.
(4) It may be ascertained even when details of extreme items are not available.

Disadvantages.

(1) A necessary preliminary to locating the mode is the arraying, classifying and grouping of the data.
(2) It may not give a representative average as it may be determined by a small number of items of similar magnitude.
(3) Its uses are limited from a mathematical point of view, e.g. if the magnitude of the mode and the number of items is known, it is not possible to calculated the aggregate of the series.
(4) It cannot be calculated by mathematical methods without preliminary processes.

11. *The Median.*

The median is the mid-most item in a series of arrayed data and is situated at the point on the x axis where the frequency of greater and smaller items is equal.

It is ascertained by—

(1) Constructing a cumulative frequency table and taking the mid-most example, i.e. the item situated at the position $\frac{n+1}{2}$.

Example 22.

Rateable Properties between £5 and £45 Rateable Value

Rateable Value	Frequency	Cumulative Frequency
Not exceeding £5	276	276
Exceeding—		
£5 and not exceeding £10 ..	843	1,119
10 ,, ,, 15 ..	1,007	2,126
15 ,, ,, 20 ..	1,984	4,110
20 ,, ,, 25 ..	3,010	7,120
25 ,, ,, 30 ..	2,476	9,596
30 ,, ,, 35 ..	1,540	11,136
35 ,, ,, 40 ..	781	11,917
40 ,, ,, 45 ..	352	12,269
	12,269	

The median is situated at the point $\frac{12{,}269 + 1}{2}$

the 6,135th item, or between £20 and £25.

(2) Constructing an ogive and taking the mid-most example. (See Figure 20, £23 10s. 0d. approximately.)

(3) Where there is no mid-most example, i.e. when there is an even number of items, or where it is situated between class limits, e.g. between £20 and £25, it may be located approximately by interpolation. The application of interpolation presupposes that the data is evenly distributed between the class limits.

The formula for interpolation is—

Let M = median.
c = class interval of the class containing the median.
L = lower limit of the class.
f = number of items in the class.
i = the number of items from the lower limits of the class to where the median occurs.

Then $M = L + \frac{c \times i}{f}$.

Example 23.

The figures are taken from the cumulative frequency table used in method 1.

$$M = 20 + \frac{5 \times 2{,}025}{3{,}010} = £23 \quad 7s.\ 6d.$$

Note.—As with the example of the mode, rateable values being calculated to the nearest £, the median would be approximated to £23.

Advantages.

(1) In many cases it can be obtained by inspection.

(2) It is easy to locate, subject to the actual number of items being known.
(3) It is usually an actual example.
(4) It is not unduly affected by abnormalities as its location depends on the number of items and not their magnitude.
(5) It admits of interpolation where data is incomplete.
(6) Its position is more definite than that of the mode.

Disadvantages.

(1) A necessary preliminary to locating the median is the arraying, classifying and grouping of the data.
(2) It may not give a representative average as—
    (1) its position is determined by the number of the items only;
    (2) it may be difficult to locate, and when located, not representative where the items in a series are not situated closely together;
    (3) its uses are limited from a mathematical point of view, i.e. the aggregate of the series cannot be calculated from the median;
    (4) it cannot be calculated by mathematical methods (subject to interpolation).

12. *Quartiles.*

The term 'Quartiles' refers to the positions one-quarter and three-quarters up an arrayed series. With the median they divide an array into four parts and are thus situated midway between the median and the extreme items. They thus give an average of a selected portion of the data, similarly in the case of deciles and percentiles, *post.*

Their location is ascertained in a similar manner to that applicable to the median, subject to the following modifications:—

(1) 1st (lower) quartile $(Q_1) = \frac{n + 1}{4}$.

(2) 3rd (upper) quartile $(Q_3) = \frac{3\ (n + 1)}{4}$.

Example 24.

In the cumulative frequency table on page 72 the 1st quartile is situated at the position $\frac{12{,}269 + 1}{4}$ the 3,067th item, i.e. between £15 and £20, and the 3rd quartile $\frac{3\ (12{,}269 + 1)}{4}$ the 9,202nd item, i.e. between £25 and £30. (See also Figure 20.)

Grouping and interpolation also may be performed as illustrated under the heading of the median.

13. *Deciles.*

The term 'deciles' refers to the positions that divide an array into tenths. Their location is ascertained in a similar manner to that applicable to the median, subject to the following modifications:—

$$\text{1st decile } \frac{n+1}{10} \ldots \text{ 6th decile } \frac{6\,(n+1)}{10}$$

Example 25.

In the cumulative frequency table on page 72 the 6th decile is situated at the position $\frac{6\ (12{,}269\ +\ 1)}{10}$, the 7,362nd item, i.e. between £25 and £30. (See also Figure 20.)

Grouping and interpolation also may be performed as illustrated under the heading of the median.

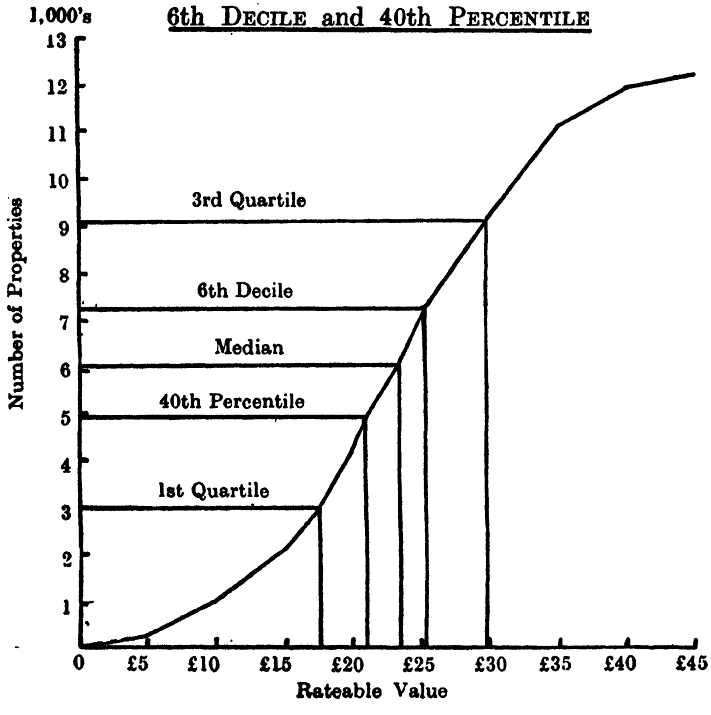
14. *Percentiles.*

The term 'percentile' refers to the positions that divide an array into hundredths. Their location is ascertained in a similar manner to that applicable to the quartiles, subject to the following modifications:—

$$\text{1st percentile } \frac{n+1}{100} \ldots \text{ 40th percentile } \frac{40\,(n+1)}{100}.$$

Figure 20.

OGIVE SHOWING MEDIAN, 1st and 3rd QUARTILES, 6th DECILE and 40th PERCENTILE



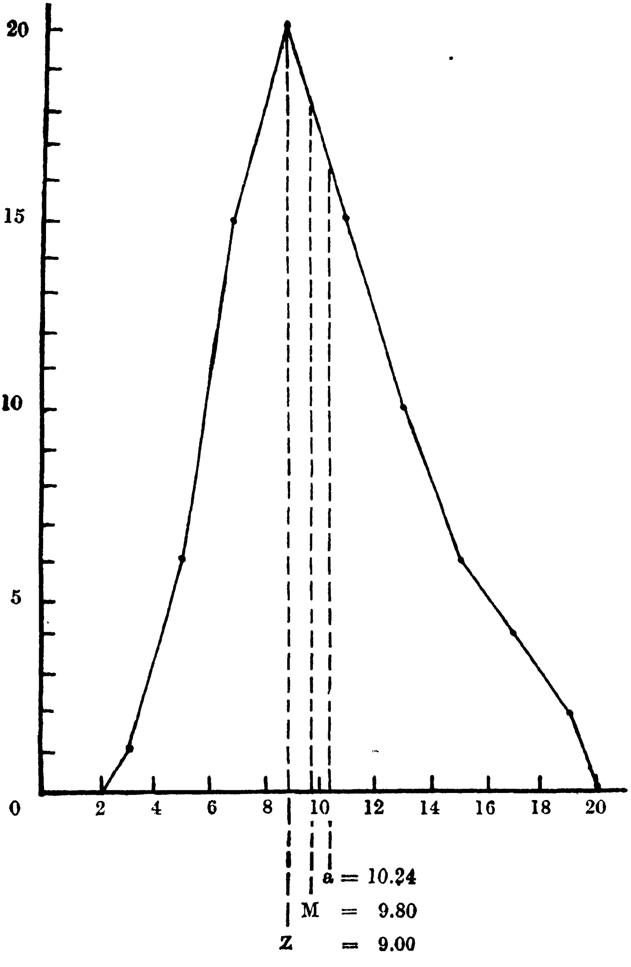
Example 26.

In the cumulative frequency table on page 72 the 40th percentile is situated at the position $\frac{40\ (12{,}269 + 1)}{100}$, the 4,908th item, i.e. between £20 and £25. (See also Figure 20.)

Grouping and interpolation may be performed as illustrated under the heading of the median.

Figure 21.

FREQUENCY POLYGON SHOWING POSITION OF MEANS



AVERAGES

A further example is given of the calculation of averages, in this case from a frequency distribution.

Example 27.

Items	Value *	Mid-Value (m)	Frequency (f)	Accumulative Frequency	fm	Log m.	$f \times$ Log m.	$\frac{1}{m}$	$\frac{f}{m}$
1	2–4	3	1	1	3	0.4771	0.4771	.333	0.333
2	4–6	5	6	7	30	0.6990	4.1940	.200	1.200
3	6–8	7	15	22	105	0.8451	12.6765	.143	2.145
4	8–10	9	20	42	180	0.9542	19.0840	.111	2.220
5	10–12	11	15	57	165	1.0414	15.6210	.091	1.365
6	12–14	13	10	67	130	1.1139	11.1390	.076	0.760
7	14–16	15	6	73	90	1.1761	7.0566	.067	0.402
8	16–18	17	4	77	68	1.2304	4.9216	.059	0.236
9	18–20	19	2	79	38	1.2788	2.5576	.053	0.106
			79		809		77.7274		8.767

*i.e. over 2 and not exceeding 4, &c.

The Arithmetic Average $= \frac{\Sigma M_w}{w}$ or $\frac{\Sigma fm}{f}$ where the weights are represented by the frequencies (f) $= \frac{809}{79}$

$$a = 10.24.$$

The Mode. (*a*) By Inspection.

Z = 9.00

(*b*) By Grouping and Interpolation.

Items				Frequencies	
1	1	7		22	
2	6		21		41
3	15	35			
4	20		35	45	
5	15	25			31
6	10		16		
7	6	10		12	
8	4		6		
9	2				

The largest frequency 35 1st grouping contains items 3 4
,, ,, ,, 35 2nd grouping ,, ,, 4 5
,, ,, ,, 45 3rd grouping ,, ,, 4 5 6
,, ,, ,, 41 4th grouping ,, ,, 2 3 4

Item 4 (8—10) appears the greatest number of times and thus contains the mode.

The mode, by interpolation $= \mathrm{L} + \dfrac{f_2 c}{f_2 + f_1} = 8 + \dfrac{15 \times 2}{15 + 15}$
$Z = 9.00$.

The Median $= \dfrac{n+1}{2} = \dfrac{80}{2}$ or the 40th item = between 8 and 10.

By Interpolation. $\mathrm{M} = \mathrm{L} + \dfrac{c \times i}{f} = 8 + \dfrac{2 \times 18}{20}$

$\mathrm{M} = 9.80$.

The Geometric Average $= \Sigma(w) \sqrt{a^{w_1} \times b^{w_2} \times c^{w_3} \ldots\ldots}$

$$\text{log geometric average} = \frac{77.7274}{79}.$$

Geometric average = 9.64.

The Harmonic Average $= \dfrac{n}{\Sigma\left(\frac{1}{m}\right)} = \dfrac{79}{8.767}$.

Harmonic average = 9.01.

The Lower Quartile $= \dfrac{n+1}{4} = \dfrac{80}{4}$ or the 20th item = between 6 and 8.

By Interpolation $\mathrm{Q}_1 = \mathrm{L} + \dfrac{c \times i}{f} = 6 + \dfrac{2 \times 13}{15}$.

$\mathrm{Q}_1 = 7.73$.

The Upper Quartile $= \dfrac{3(n+1)}{4} = \dfrac{240}{4}$ or the 60th item = between 12 and 14.

By Interpolation $\mathrm{Q}_3 = \mathrm{L} + \dfrac{c \times i}{f} = 12 + \dfrac{2 \times 3}{10}$.

$\mathrm{Q}_3 = 12.60$.

The positions of the arithmetic average, the median and the mode, are indicated in Figure 21, where the data is plotted in the form of a frequency polygon.

General Caution in the Use of Averages.

A certain amount of caution is necessary in comparing averages as the data from which they are derived may not be truly homogeneous throughout the period of the inquiry.

Thus a ratio may change by reason of the fact that the groups of data from which it is derived have changed in constitution.

It is therefore necessary to ascertain whether an average is affected by—

(*a*) a change in the character of the data, as a whole, from which it is derived, or

(*b*) a change in the constitution of the individual groups of the data.

In the case of item (*a*) it may be necessary to supplement the result by further inquiries, or to alter the basis of the investigation.

In the case of item (*b*) the data should be referred to a factor designed to standardise the basis of the data over the full period of the inquiry, but to leave the variation under consideration free to operate.

For example, in the case of vital statistics—although the principles hold good for other departments of inquiry—it may not be correct to assume that mortality in two districts, or in the same district at different times, is the same, because the death rates (crude) are similar. It is known that mortality is greater in the case of old people and infants under one year of age than in the case of the intermediate age-groups. Sex also has a bearing on the subject as well as questions of environment and local or racial characteristics.

The method of dealing with these factors is to split the whole data into its constituent parts and relate each part to a representative (standardised) population age-sex distribution. Thus the local rates in the different age-sex groups are related to notional and homogeneous groups derived from the total population of the country.

The base is fixed and the age-sex element standardised and, by applying the local variable, comparable and homogeneous ratios and a combined ratio are obtained, which give effect to the local characteristics.

(The question of correction factors is discussed further in Chapter XII under the heading of 'Vital Statistics'.)

Chapter VII

TREND

Trend: Types of Fluctuations: Investigation of the Trend and Short Period and other Fluctuations.

THE undermentioned factors arise, in whole or in part, on consideration of a statistical series, extending over a period of time, viz.:—

(1) *The Trend*—i.e. the general movement of the data as distinct from temporary or short period fluctuations and irregularities or abnormalities.

(2) *Cyclical Fluctuations*—i.e. defined tendencies, of the nature of trade cycles.

(3) *Seasonal Variations*—i.e. phenomena recurring periodically, e.g. the influence of harvests.

(4) *Catastrophic Movements*—i.e. unusual events, e.g. the occurrence of a cyclone or the outbreak of war.

(5) *Other unforseen or unusual Influences*, e.g. a change in the monetary policy of the country.

Thus in many cases of statistical inquiry it is desired to investigate not the actual facts, but—

(1) The trend, which may be either
  (*a*) secular, i.e. the movement of the data over a long period of years, shorn of temporary fluctuations;
  (*b*) periodic, i.e. the general tendency.
(2) Short term fluctuations.
  (*a*) Regular (caused by periodic influences).
  (*b*) Irregular (caused by random influences).
(3) The relation of short term fluctuations to the trend.

The following operations would provide information in such an inquiry:—

Example 28.

DATA

Year		Average Price	Average Price (3 years)	Fluctuations from 3-year average price
1928	..	5s 0d	—	—
1929	..	6 0	5s 10d	+ 2d
1930	..	6 6	6 1	+ 5
1931	..	5 9	5 11	— 2
1932	..	5 6	5 4	+ 2
1933	..	4 9	4 10	— 1
1934	..	4 3	4 6	— 3
1935	..	4 6	4 5	+ 1
1936	..	4 6	—	—

Note.—The figure for the 3-year average price appears in the table opposite the centre figure of the period of the average, i.e. the average for 1928, 1929 and 1930 is shown at 1929. It is similarly shown in the graph, Figure 22.

(1) *The actual Data.*—Plot the actual data as a graph. (Historigram.) Example, Figure 22 (continuous line).

(2) *The Trend.*—Calculate a moving average and plot the resulting figures as a graph. Example, Figure 22 (dotted lines).

No definite guide can be given as to the period of the moving average. It will depend on the nature of the subject of the inquiry and the influence and magnitude of the short term fluctuations. The period should not be so lengthy, for example, as to defeat the object of the inquiry by producing data possessing very little characteristics, but should be such as to adequately show the general movement of the data, and to take account of, but not be unduly influenced by, the temporary movements. Where, for example, it is desired to eliminate the influence of a trade cycle, the period of the moving average should correspond with the period of the cycle.

Too short a period does not eliminate temporary fluctuations. Too long a period also does not eliminate these fluctuations, but their phase may be reversed.

The effect of utilising a moving average, on the trend, and on fluctuations from the trend is summarised below:—

Trend.

(*a*) Linear. Reproduces trend exactly.

(*b*) Curved:—

1. Convex to horizontal (time) axis.
   Moving average curve is above the original.
2. Concave to horizontal (time) axis.
   Moving average curve is below the original.

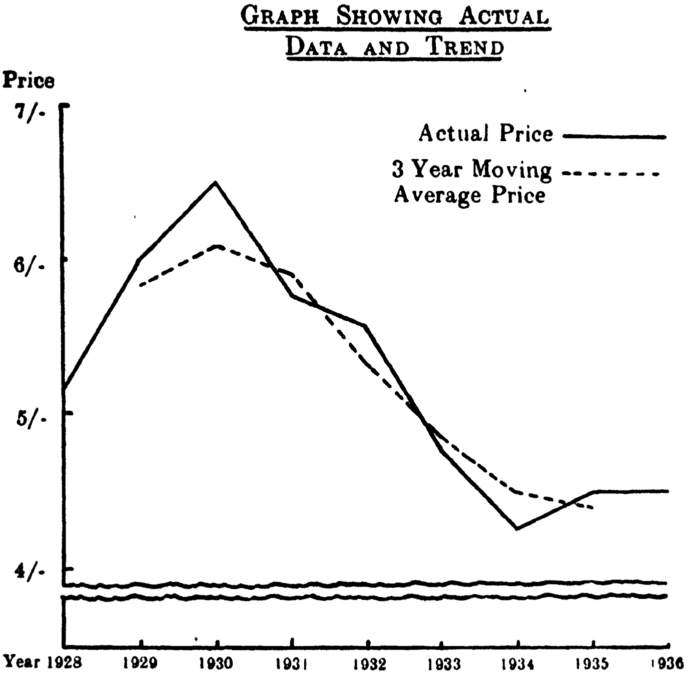
Fluctuations.

(*a*) Regular. If the number of items agrees with the period of the fluctuations, the latter are completely eliminated. With other groupings the fluctuations are reduced.

(*b*) Irregular. The fluctuations are reduced in proportion to the number of items in the moving average.

To overcome these objections, the drawing of a curve, freehand, through the observations, has much to commend it. The curve is designed to ignore minor fluctuations, and to follow the general trend of the data. This method, however, demands a high degree of skill and judgment in its application.

Figure 22.



(3) *The Short Term Fluctuations.*—Utilise the moving average as the value O on a graph and plot the short term fluctuations (+ and —) therefrom. (Alternatively the fluctuations from the arithmetic average could be utilised, but only when the general trend is steady or has a slight upward or downward slope.) Example, Figure 23.

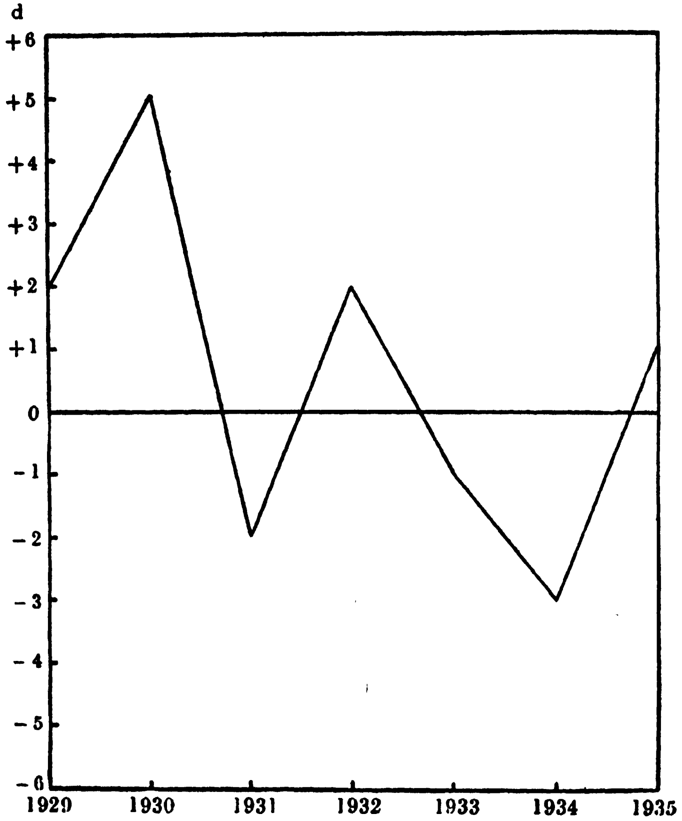
The information illustrated by the foregoing operations may be supplemented by—

(1) Plotting the data as a graph designed to show the ratio of change, i.e. by utilising a logarithmic scale.

(2) Reducing the data to index numbers (Chapter VIII) before plotting and subsequent analysis.

Figure 23.

GRAPH SHOWING FLUCTUATIONS FROM TREND



A further and more extensive example of the analysis of a time series is given below. The figures are hypothetical and are designed to serve the purpose of illustrating the successive stages of the methods that could be adopted, without being too accommodating on the one hand or too lacking in characteristics on the other.

Example 29.

DATA

Period (1)	Index Number (2)	7-period moving average (3)	Variations from moving average (4)
1	70	—	—
2	73	—	—
3	76	—	—
4	80	74.6	+ 5.4
5	76	74.9	+ 1.1
6	74	75.1	— 1.1
7	73	75.3	— 2.3
8	72	75.6	— 3.6
9	75	76.1	— 1.1
10	77	76.7	+ 0.3
11	82	77.0	+ 5.0
12	80	77.1	+ 2.9
13	78	77.3	+ 0.7
14	75	77.7	— 2.7
15	73	78.0	— 5.0
16	76	78.0	— 2.0
17	80	78.1	+ 1.9
18	84	78.6	+ 5.4
19	80	—	—
20	79	—	—
21	78	—	—

The data (Example 29) shows index numbers (column 2) for a period of 21 years. From an examination of the numbers it appears that peaks occur in the case of the years numbered 4, 11 and 18, and troughs in respect of years 1, 8 and 15. These facts are even more evident when the data is plotted as a graph (Figure 24, continuous line). Further, the whole series and supporting curve indicate a slight, though definite, increase in values over the whole period. It is therefore reasonable to assume that the series is subject to a cycle of about 7 years, with a long term trend tending to a rise in values.

On calculation, a 7-year moving average (column 3) shows a steady increase from 74.6 to 78.6. These averages, when plotted, form a reasonably straight line, indicating that short term fluctuations have largely been eliminated. (Figure 24, broken line.)

The fluctuations from the moving average, however, are somewhat erratic (column 4), and when plotted with the trend as zero (Figure 25) only exhibit a tendency to follow the period of the 7-year cycle.

Further details, therefore, are required, and these are given in the form of quarterly index numbers for a portion of the original table.

Example 30.

DATA

Period (1)		Index Number .. (2)	Sum in 4's (3)	Addition in pairs (4)	Division by 8 (Trend) (5)	Total Fluctuations (6)	Periodic Fluctuations (7)	Random Fluctuations (8)
4	1	81						
	2	77						
			320					
	3	79		636	79.5	— .5	— .8	+ .3
			316					
	4	83		628	78.5	+ 4.5	+ 3.0	+ 1.5
			312					
5	1	77		620	77.5	— .5	+ .7	— 1.2
			308					
	2	73		612	76.5	— 3.5	— 2.9	— .6
			304					
	3	75		606	75.8	— .8	— .8	0
			302					
	4	79		602	75.0	+ 4.0	+ 3.0	+ 1.0
			300					
6	1	75		598	74.8	+ .2	+ .7	— .5
			298					
	2	71		594	74.2	— 3.2	— 2.9	— .3
			296					
	3	73		591	73.9	— .9	— .8	— .1
			295					
	4	77		589	73.6	+ 3.4	+ 3.0	+ .4
			294					
7	1	74		587	73.3	+ .7	+ .7	0
			293					
	2	70		585	73.1	— 3.1	— 2.9	— .2
			292					
	3	72		583	72.9	— .9	— .8	— .1
			291					
	4	76		581	72.6	+ 3.4	+ 3.0	+ .4
			290					
8	1	73		579	72.4	+ .6	+ .7	— .1
			289					
	2	69		577	72.1	— 3.1	— 2.9	— .2
			288					
	3	71		579	72.4	— 1.4	— .8	— .6
			291					
	4	75		585	73.1	+ 1.9	+ 3.0	— 1.1
			294					
9	1	76		593	74.1	+ 1.9	+ .7	+ 1.2
			299					
	2	72		601	75.0	— 3.0	— 2.9	— .1
			302					
	3	76		604	75.5	+ .5	— .8	+ 1.3
			302					
	4	78		606	75.8	+ 2.2	+ 3.0	— .8
			304					
10	1	76		608	76.0	0	+ .7	— .7
			304					
	2	74		610	76.2	— 2.2	— 2.9	+ .7
			306					
	3	76		619	77.4	— 1.4	— .8	— .6
			313					
	4	80		631	78.9	+ 1.1	+ 3.0	— 1.9
			318					
11	1	83		641	80.1	+ 2.9	+ .7	+ 2.2
			323					
	2	79		651	81.4	— 2.4	— 2.9	+ .5
			328					
	3	81		654	81.8	— .8	— .8	0
			326					
	4	85		650	81.2	+ 3.8	+ 3.0	+ .8
			324					
12	1	81		646	80.8	+ .2	+ .7	— .5
			322					
	2	77		642	80.2	— 3.2	— 2.9	— .3
			320					
	3	79						
	4	83						

The object now is to isolate the periodic or seasonal fluctuations.

The first step is achieved by calculating the general trend. The series is totalled in periods of four and these sums set midway between the first and fourth periods of each average, e.g. in the first item—midway between the second and third quarterly periods (column 3).

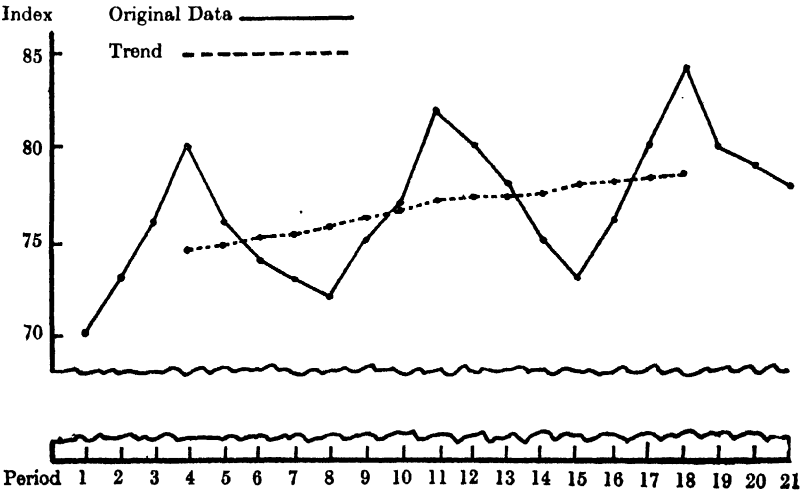
These sums of four quarterly periods are then totalled in pairs (column 4). Division by 8 gives the general trend; which is set opposite the third item in each group taken to form the average, e.g. in the first item—opposite the third quarterly period (column 5). The effect of this method is to set the trend opposite, instead of between, items in the quarterly periods. It also increases the range of the moving average and thus provides more stability to the trend.

Figure 26 shows the actual data (continuous line) and the trend (broken line).

The next step is to ascertain the average quarterly or seasonal variation. This is achieved by setting out the fluctuations from the trend (column 6) and obtaining a quarterly average.

Figure 24.

Graph Showing Actual Data and Trend



Example 31.

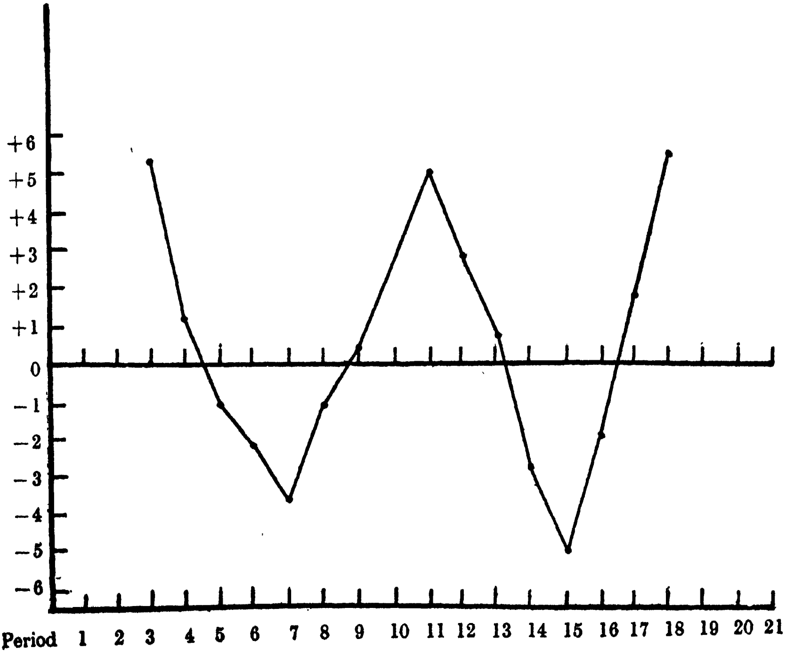
Data as in Example 30, column 6 (total fluctuations).

Period	1.	2.	3.	4.
4			— .5	+ 4.5
5	— .5	— 3.5	— .8	+ 4.0
6	+ .2	— 3.2	— .9	+ 3.4
7	+ .7	— 3.1	— .9	+ 3.4
8	+ .6	— 3.1	— 1.4	+ 1.9
9	+ 1.9	— 3.0	+ .5	+ 2.2
10	0	— 2.2	— 1.4	+ 1.1
11	+ 2.9	— 2.4	— .8	+ 3.8
12	+ .2	— 3.2		
	8)+ 6.0	—23.7	— 6.2	+24.3
Seasonal average	+ .7	— 2.9	— .8	+ 3.0

The average quarterly or seasonal variation is set out in Example 30, column 7, and plotted in Figure 27.

Figure 25.

GRAPH SHOWING FLUCTUATIONS FROM TREND



This method of calculating the quarterly average is based on the assumption that the random fluctuations will be fairly stable on the average when calculated over a reasonably long period. Thus random plus values will tend to offset minus values, leaving the trend unaffected. It is reasonable to presume that prior knowledge would be had of occurrences of a catastrophic nature, when allowance for items of this character should be made when calculating the average seasonal variation. The quarterly or seasonal average (Example 30—column 7) is then deducted from the total fluctuations (column 6), leaving the random short term fluctuations (column 8). These do not normally exhibit any particular or regular characteristics (Figure 28, dotted line), and would thus form the basis for further investigation of the original data as to their cause.

Figure 26.

GRAPH SHOWING ACTUAL DATA

AND SHORT TERM TREND

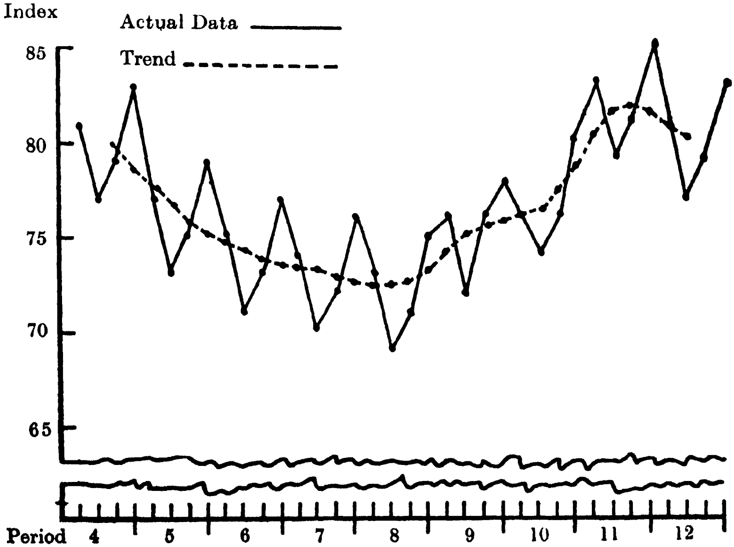


Figure 27.

GRAPH SHOWING TOTAL
SHORT TERM FLUCTUATIONS FROM TREND

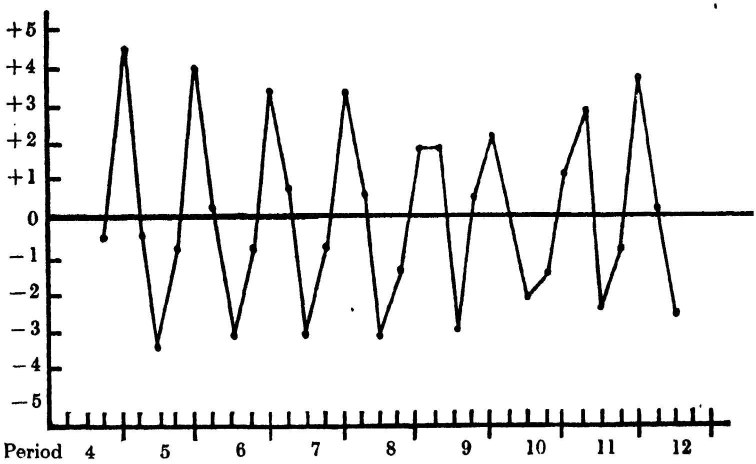
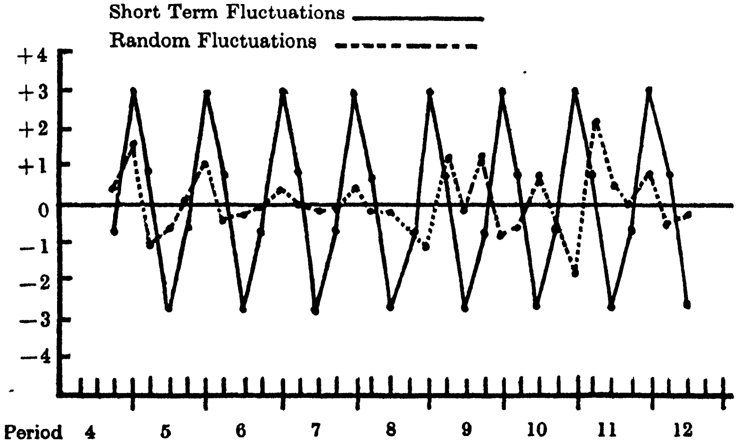


Figure 28.

GRAPH SHOWING SHORT TERM AND RANDOM
FLUCTUATIONS FROM TREND



Chapter VIII

INDEX NUMBERS

Nature: Principles: Factors: Cautions in Use: Construction: General Principles: Fixed Base Method: Simple Index Numbers: Weighted Index Numbers: Chain Base Method: Published Index Numbers.

INDEX numbers are numerical devices prepared to show the relation existing between facts at different points of time. They are thus a measurable characteristic expressed as a proportion of a standard value of that characteristic or a special form of average used to indicate changes and to measure the relative movements of variables.

The principle of index numbers is that all factors to be compared are reduced to a common denominator.

The facts at a certain time, or an average, are used as the basis of the calculation (termed the 'base') to which the new factors are expressed as ratios.

Index numbers may be prepared in respect of individual items and composite index numbers in respect of groups of items.

Generally speaking, index numbers are constructed to show the net effect of the subjects in general and not the movements of any particular item. They thus form an indication of the trend or movement of the data as a whole. They may also be so constructed as to enable variations in particular items to be compared or to form the basis for investigation into the causes of fluctuations.

To construct an index number from the whole of the data available is, in the majority of cases, impracticable. For example, the task of collecting and preparing complete data for calculating an index representative of the cost of living would be enormous, if not impossible. In such cases recourse is had to sampling. The selection of suitable samples and the weights, if any, to be given to them is an extremely complicated and technical matter presenting special problems varying with the subject under review.

Construction.—The phases in the construction of an index number are:—

(1) The selection of the items.
(2) The selection of the base period.
(3) The periodical ascertainment of the values of the items.
(4) The combination of the values to form a composite index number.

Note.—In order to avoid undue generalisations the illustrations and notes following are given relative to an index depicting the cost of living at different dates.

(1) The selection of the items.

(*a*) The samples selected should be of the same relative importance at different times.

(*b*) Samples should be homogeneous throughout the currency of the index number.

(*c*) Samples should be so chosen that they are truly representative of the subject as a whole.

(*d*) Movements in one item should not be caused by movements in another, otherwise tendencies may be negatived or given undue prominence; subject, however, to the proviso that the selection of samples and the weights allotted to them may have been undertaken with this factor in mind.

(*e*) The price to be utilised, viz., wholesale or retail. Wholesale prices are more uniform throughout the country. Where retail prices are selected, the data should be the average of a number of prices obtained from representative parts of the country.

(2) The selection of base period.

(*a*) This may conveniently be the prices ruling in a certain representative year or an average of the prices ruling in a number of years.

(*b*) The base, however devised, should represent a period of normal conditions free from abnormalities.

(3) The periodical ascertainment of the values of the items.
Samples should be selected, the prices and quantities of which can be accurately and easily obtained.

(4) The combination of the values to form a composite index number.

(*a*) The items in the base period selected are given the value of 100 and the sum of the indices so obtained are equated to 100, giving the basic index number.

(*b*) The ratio (percentage) of the increase or decrease in subsequent periods, termed the 'price relative', is ascertained in respect of each of the individual items by dividing the price relating to the period under review by the base period, and multiplying this result by 100 (e.g. Commodity A in Example 32, $\frac{3\frac{1}{2}}{2} \times 100 = 175$).

(*c*) The arithmetic or geometric mean of the sum of the indices or 'proportional numbers' so obtained is then extracted. This calculation gives the index number for the period under review.

In abstract theory it is permissible to employ any form of average in arriving at the composite index number. In practice it is found that the only averages suitable or practicable are the arithmetic and geometric means, although in some instances the harmonic mean is used.

Where, however, individual items used in the calculation of the final index number are subject to a rising (or falling) trend, the method of taking a simple or weighted arithmetic average has the effect of increasing (or decreasing) the weight applicable to those items. This disadvantage does not arise with the use of the geometric average.

The chief arguments in favour of the use of these averages are:—

(1) *The Arithmetic Mean.*
  (*a*) It is simple to calculate.
  (*b*) Its significance is generally understood.

(2) *The Geometric Mean.*
  (*a*) Owing to its nature it is more suitable than the arithmetic average, especially in the case of the chain base method of calculation (*post*). It measures the ratio of change (the relative difference) and is thus appropriate, especially when, as is the case in index numbers, an average of percentages is required.
  (*b*) It is smaller than the arithmetic average. Arising from this fact, it is asserted that as a rise in price usually produces a tendency for decreased purchases, the geometric mean provides a useful corrective factor in the resulting composite index number.

The possibilities of the principle of index numbers are great, but extreme caution is necessary in deducing conclusions from them, owing to their principles of construction. The necessity for the revision of the basis or for allowing for factors that may arise during the course of time are illustrated by the following points, viz.:—

(1) During the course of time the items may not be homogeneous, e.g. the food value of a particular item may be greatly increased by scientific methods of preparation.

(2) The introduction of more efficient production or distribution may seriously affect prices.

(3) The basis may not be representative owing to the possibility of

'new wants', in the economic sense, arising or supplanting original factors in the basis.

(4) Too much reliance cannot be placed upon past standards owing to the progress of civilisation and the general standard of living.

The methods of the construction of index numbers are illustrated below.

Example 32.

1. *Simple Index Numbers* (Fixed base method).

	1926			1936		
Commodity	Price		Index No.	Price		Index No.
A ..	2d.	=	100	3½d.	=	175
B ..	1s. 0d.	=	100	1s. 6d.	=	150
C ..	8d.	=	100	2d.	=	25
D ..	1s. 8d.	=	100	1s. 4d.	=	80
E ..	4d.	=	100	8d.	=	200
			5)500			5)630
Index Numbers			100			126

Formula.

Let P_B = the price at the base period.

P_C = the price at the period next following

then the Price relative $= \dfrac{P_C}{P_B}$

and the index number of the commodity

$$= \frac{P_C}{P_B} \times 100.$$

Then the composite index number

(*a*) When individual commodity numbers are required

$$= \frac{\Sigma\left(\dfrac{P_C}{P_B} \times 100\right)}{n}$$

(*b*) When a general index only is required

$$= \frac{\Sigma \dfrac{P_C}{P_B}}{n} \times 100$$

The procedure where it is necessary to introduce a new or discard an old item, is to calculate the index under both sets of circumstances. The new index is then equated to the old by means of a factor derived from the indices for the period of transition.

Example **33.**

Period	Old Index	New Index
1	110	
2	115	
3	110	100
4		105
5		108
6		112

The equation figure, derived from the indices calculated under both old and new circumstances for period No. 3 (the date of change) is $\frac{110}{100} = 1.1$.

By multiplying the new index by this factor, comparable figures are obtained for the full period of the index.

The resulting series of index numbers are as follows:—

	Index			
Period	Old	New		
1	110			
2	115			
3	110	110	i.e.	(100 × 1.1)
4		115.5		(105 × 1.1)
5		118.8		(108 × 1.1)
6		123.2		(112 × 1.1)

2. *Weighted Index Numbers* (Fixed base method). The example given (simple index numbers) assumes that the relative values of the commodities are equal, or that an equal amount of each commodity is demanded, and should this be so, the index number will be of practical use. In an index number based on a large number of carefully selected items, it is possible that a basis could be devised to be truly representative of the subject under consideration. If, however, the items are of varying importance and they have not been selected so as to offset tendencies in one direction by tendencies in another, recourse must be made to weights in calculating the index number. The relative values or quantities of the commodities A, B, C, D, E in the illustration may be 4, 1, 4, 2, 3, respectively. These factors form the weights to be used in the calculation.

In introducing weighting factors the indices (unweighted), price relatives or prices, are multiplied by the weight applicable, and the total of the products so obtained are divided by the total of the weights.

Formula.

Let w = the weight applicable.

Then (*a*) when weights are applied to the price relative (the average of ratios method), the index number $= \dfrac{\Sigma \dfrac{P_C}{P_B} \times w}{\Sigma w} \times 100.$

(*b*) when weights are applied to the prices (the aggregative method), the index number $= \dfrac{\Sigma P_C \times w}{\Sigma P_B \times w} \times 100.$

Note.—The undermentioned systems of weighting have been adopted in practice:—

1. Actual quantities of the base year.
2. Actual quantities of the current year.
3. Fixed weights designed to represent a typical year.
4. A hybrid or crossed system making use of a combination of the foregoing methods.

For Example:—The sum and mean of:—

1. $\dfrac{\text{Base year quantities at current prices}}{\text{Base year quantities at base prices}}$
2. $\dfrac{\text{Current year quantities at current prices}}{\text{Current year quantities at base prices}}$
3. $\dfrac{\text{Fixed quantities at current prices}}{\text{Fixed quantities at base prices}}$
4. The square root of $\dfrac{\text{Base year quantities at current prices}}{\text{Base year quantities at base prices}} \times \dfrac{\text{Current year quantities at current prices}}{\text{Current year quantities at base prices}}$
5. $\dfrac{(\text{Base year quantities} + \text{current year quantities})}{(\text{Base year quantities} + \text{current year quantities})} \times \dfrac{\text{Current prices}}{\text{Base prices}}$

Example 34.

Weighted Index Number

Commodity	Weights	Indices	Product
A	4	175	700
B	1	150	150
C	4	25	100
D	2	80	160
E	3	200	600
	14		14)1,710
	Index Number (weighted)		122

3. *Chain Base Method.*—The examples given are calculated on the fixed base method. The term is self explanatory and refers to the fact that the base, when decided upon, remains unaltered throughout the period of the calculations, comparison thus being made between the base period and ensuing periods. This feature may not always be desirable, so where applicable, a further method—the chain base method—is used. In this method the changes in price are measured by the figures applicable to the previous year, or other suitable period, and not to a fixed base period. In other respects the calculation is the same. A moving or chain base enables certain objections raised against fixed base index numbers to be negatived, e.g. :—

(*a*) it may be necessary to make comparisons in the numbers from year to year rather than to a base year—which may be remote, especially is this so when considering short term fluctuations;

(*b*) the chain base method is more suitable than the fixed base method when it is desired to introduce new factors, or to discard old factors, in the base of the number.

An example of the effect of applying the two methods is given below. For the sake of simplicity the question of weighting has been ignored. The index number under both methods is the same in the first year (1935) as the base is the same.

Formula.

$$\text{Index Number, second year} = \frac{\Sigma \dfrac{P_C}{P_B}}{n} \times 100.$$

$$\text{Index Number, third year} = \frac{\Sigma \dfrac{P_D}{P_C}}{n} \times 100.$$

Example 35.

Comparison of Fixed and Chain Base Methods.

DATA

Commodity	Price 1934 s	Price 1934 d	Price 1935 s	Price 1935 d	Price 1936 s	Price 1936 d
A		6		8		7
B		4		3½		4
C	1	1	1	0	1	2
D		9		11		10
E	2	3	2	7	2	4

Commodity	Base 1934	Indices 1935 (Fixed & Chain Base)	Indices 1936 (Fixed Base)	Indices 1936 (Chain Base)
A	100	133	117	88
B	100	88	100	114
C	100	92	108	117
D	100	122	111	91
E	100	115	104	90
	5)500	5)550	5)540	5)500
Index Numbers	100	110	108	100

When required, a chain base index may be converted to a fixed base. As the indices are the same for the base and first periods, this conversion is effected by multiplying the fixed base index for the first period by the chain base index for the next ensuing period, and dividing the result by 100.

Example 36.

Period	Chain Base Index	Fixed Base Index	
—	—	—	
Base	100	100	
1	105.5	105.5	
2	105.3	111.1	i.e. $\frac{105.5 \times 105.3}{100}$
3	95	105.5	i.e. $\frac{111.1 \times 95}{100}$

Continued on page 97

Continued from page 96.

Period	Chain Base Index	Fixed Base Index	
—	—	—	
4	110.5	116.6	i.e. $\frac{105.5 \times 110.5}{100}$
5	104.8	122.2	i.e. $\frac{116.6 \times 104.8}{100}$

Note.—The basic figures for the above indices are as follows:—

Period	Basic Figures
—	—
Base	90
1	95
2	100
3	95
4	105
5	110

Notes on the chief index numbers published are given below:—

The material in regard to item No. 1, the Ministry of Labour Cost of Living Index Number, is set out in greater detail so as to afford some indication of the practical difficulties and problems involved in the preparation of index numbers. The remaining notes are in the form of summaries, viz.:—

1. *Cost of Living Index Number* (Ministry of Labour).

(*a*) *Use.*—To measure the average increase in the cost of maintaining unchanged the pre-war (1914) standard of living of the working classes. It is also used by itself or in combination with a wholesale price index as a relative measure of changes in the purchasing power of money. It is also extensively utilised in adjusting basic rates of wages, often by means of a sliding scale depending on the state of the percentage increase (or decrease) of the index.

(*b*) *Base Period.*—1914 = 100.

(*c*) *Base.*—The following groups:—
   (*a*) Food.
   (*b*) Rent.
   (*c*) Clothing.
   (*d*) Fuel and Light.
   (*e*) Sundries.

(*d*) *Method of Construction.*

(1) Information is obtained in regard to:—

(*a*) *Food*, by inquiries addressed to all types of retailers enjoying a working-class trade. Over 5,000 retailers in over 500 towns and villages are approached.

From this information the average percentage increase in price over those ruling in July, 1914, is calculated for each item of food. These percentages are weighted by figures based on the average expenditure of 1,944 working class families. The latter were obtained as the result of an investigation undertaken by the Board of Trade in 1904.

The items of food and weights employed are as follows:

Beef	48	Sugar	19
Mutton	24	Milk	25
Bacon	19	Butter	41
Fish	9	Cheese	10
Flour	20	Margarine	10
Bread	50	Eggs	19
Tea	22	Potatoes	18

The weighted percentages are combined and result in a weighted average increase (percentage) over the prices ruling in 1914.

(*b*) *Rent*, by inquiries addressed to local authorities and property owners' associations. The information covers both controlled and decontrolled dwellings. The final figure is given in the form of an average percentage increase for the whole country.

(*c*) *Clothing*, by inquiries addressed to outfitters, drapers, boot retailers, &c. Over 300 retailers in 81 towns are approached and the inquiry covers all types and quality of clothing purchased by the working classes. Price relatives are calculated on the chain base principle, and combined into a weighted average, which allows for making garments in cases in which materials are purchased.

(*d*) *Fuel and Light*, by inquiries addressed to coal merchants, gas undertakings and retailers. The inquiry covers coal, gas, oil, candles and matches. The percentage increases are combined into a weighted average calculated upon the fixed base method.

(*e*) *Other Items*, by inquiries addressed to retailers, transport undertakings and the Press. The inquiry covers soap, soda, ironmongery, brush-ware, pottery, tobacco, cigarettes, fares and newspapers. The percentage increases are combined into a weighted average calculated upon the fixed base method.

(2) Retail prices are utilised.

(3) Weights are utilised for each group (based on a survey of working class family budgets, carried out in 1904 by the Board of Trade, an inquiry into rents in 1912 and other available information), and are designed to adjust the factors from the aspect of their relative importance. These weights are as follows:—

(*a*) Food	7½
(*b*) Rent	2
(*c*) Clothing	1½
(*d*) Fuel and Light ..	1
(*e*) Other Items ..	½
	12½

(4) The resulting number is expressed in the form of an increase in the cost of living. To convert this figure into an index number it is necessary to add 100.

2. *Index of Industrial Production* (Board of Trade).

(*a*) *Use.*—To measure the value added in manufacture or production to the materials used.

(*b*) *Base Period.*—1930 = 100.

(*c*) *Base.*—60 branches of industry.

(*d*) *Method of Construction.*

(1) Information is obtained from—

(*a*) The Census of Production.

(*b*) The official returns of imports and exports, of wages paid, and of production.

(*c*) Information supplied by trade associations, industrial federations, private firms, and trade newspapers, showing production and movements in stocks.

(2) Grouping is adopted as follows:—

(1) Mines and quarries
(2) Iron and steel
(3) Non-ferrous metals
(4) Building materials and building
(5) Engineering and ship building
(6) Textiles
(7) Chemicals and oils

- (8) Leather, boots and shoes
- (9) Food, drink and tobacco
- (10) Gas and electricity

(3) Individual indices are weighted in proportion to the net output, i.e. the value added in course of manufacture or production to the materials used.

3. *Index of Average Prices* (Sauerbeck).
   - (*a*) *Use.*—To estimate the level of wholesale prices.
   - (*b*) *Base Period.*—The arithmetic average of the years 1867-77 = 100.
   - (*c*) *Base.*—45 commodities.
   - (*d*) *Method of Construction.*
     - (1) The chief sources of information are the Board of Trade returns.
       The average prices of commodities are obtained by dividing the total value of imports or exports of particular commodities by the total units imported or exported.
     - (2) No methods of grouping or weighting, are utilised.
     - (3) The arithmetic average of the indices forms the index number.

4. *The* 'Statist' *Index Number.*
   - (*a*) *Use.*—To estimate the level of wholesale prices.
   - (*b*) *Base Period.*—The arithmetic average of the years 1867-77 = 100.
   - (*c*) *Base.*—45 commodities.
   - (*d*) *Method of Construction.*
     1. The index number is a continuation of the 'Sauerbeck' number.
     2. The items are grouped as follows:—
        - (1) *Foods.*
          - (*a*) Vegetable, 8 commodities.
          - (*b*) Animal, 7 commodities.
          - (*c*) Sugar, coffee and tea, 4 commodities.
            Total, 19 commodities.
        - (2) *Materials.*
          - (*a*) Minerals, 7 commodities.
          - (*b*) Textiles, 8 commodities.
          - (*c*) Miscellaneous, 11 commodities.
            Total, 26 commodities.
        - (3) A separate index number is calculated for each group and also for each class.

(4) The general index is the simple arithmetic mean of the 45 items.

5. *The* 'Economist' *Index Number.*
   - (*a*) *Use.*—To estimate the level of wholesale prices.
   - (*b*) *Base Period.*—The average prices of the year 1927 = 100.
   - (*c*) *Base.*—58 commodities.
   - (*d*) *Method of Construction.*
     - (1) The sources of information are the Board of Trade returns.
     - (2) The items are grouped as follows:—
       - (*a*) Cereals and meat.
       - (*b*) Foods, other than cereals.
       - (*c*) Textiles.
       - (*d*) Minerals.
       - (*e*) Miscellaneous.
     - (3) Weighting is resorted to in the case of certain commodities.
     - (4) A separate index number is calculated for each group.
     - (5) The composite index number is calculated by taking the geometric mean of the index numbers obtained from the 58 commodities.

6. *The 'Board of Trade' Index Number.*
   - (*a*) *Use.*—To estimate the level of wholesale prices.
   - (*b*) *Base Period.*—The corresponding month of the previous year, commencing at 1930.
   - (*c*) *Base.*—200 commodities or classes.
   - (*d*) *Method of Construction.*
     - (1) Market values (wholesale) form the basis of the prices used.
     - (2) The commodities are grouped as follows:—

(1) Foods.

(*a*) Cereals	20	Items
(*b*) Meat, fish and eggs	20	,,
(*c*) Miscellaneous foods and tobacco	28	,,
Total	68	

(2) Other Items.

(*a*) Coal	9	Items
(*b*) Iron and Steel	37	,,
(*c*) Non-ferrous metals	8	,,
(*d*) Cotton	10	,,
(*e*) Wool	11	,,
(*f*) Other textiles	9	,,
(*g*) Chemicals and oils	15	,,
(*h*) Miscellaneous	33	,,
Total	132	

(3) Weighting in the ordinary sense of the term is not resorted to, but an equivalent effect is obtained by introducing additional quotations of items of importance, as indicated by the Census of Production.

(4) The chain base method is utilised in the calculation, the basis being the corresponding figure of the previous year.

(5) A separate index number is calculated for each of the 200 commodities, and also for each class and group.

(6) The composite index number is calculated by taking the geometric mean of the index numbers obtained from the 200 commodities.

7. *The* 'Financial Times' *Index Number.*

(*a*) *Use.*—To estimate the level of wholesale prices.

(*b*) *Base Period.*—The average prices ruling in the year 1920.

(*c*) *Base.*—73 commodities.

(*d*) *Method of Construction.*

(1) Prices are obtained from commercial reports of newspapers, trade journals, &c.

(2) The commodities are grouped as follows:—

(*a*)	Cereals	7	Items
	Meat	8	,,
	Miscellaneous foods	8	,,
	Total	23	
(*b*)	Cotton	8	Items
	Textiles	9	,,
	Iron and steel	12	,,
	Other metals	5	,,
	Fuel	5	,,
	Miscellaneous	11	,,
	Total	50	

(3) The chain base method is utilised in the calculation.

(4) The composite index number is obtained by calculating the geometric mean of the various items.

(5) A weekly figure is prepared and the monthly index figure is obtained by calculating the arithmetic average of the weekly numbers.

8. *Index of Business Activity* (*The* 'Economist').
   - (*a*) *Use.*—To show the trend of trade quantitatively and not in terms of values. It thus attempts to measure variations in the 'real' national income.
   - (*b*) *Base Period.*—1924 = 100.
   - (*c*) *Base.*—The following groups:—

	Weights
(1) Employment	10
(2) Consumption of Coal	4
(3) Consumption of Electricity	2
(4) Railway merchandise traffic	4
(5) Commercial vehicles in use	2
(6) Postal traffic recipts	3
(7) Building activity	2
(8) Consumption of iron and steel	2
(9) Consumption of cotton	1
(10) Imports of raw materials	2
(11) Exports of British manufactures	3
(12) Movements of shipping	2
(13) Metropolitan, Country and Provincial Bank clearings	4
(14) Town Bank clearings	1

   - (*d*) *Method of Construction.*
     - (*a*) Each group is weighted to allow for—
       - (1) Size of industry.
       - (2) Its significance as an indicator of business activity.
       - (3) Its speed of development or its maturity.
       - (4) Its steadiness or liability to fluctuation.
       - (5) Seasonal fluctuations are removed by correction factors.
     - (*b*) The composite index number is obtained by calculating the geometric mean of the various groups.

9. *Index of Aggregate Profits* (Lord Stamp).
   - (*a*) *Use.*—To estimate the aggregate profits of industry.
   - (*b*) *Base Period.*—1924 = 100.
   - (*c*) *Base.*
     - (*a*) Income-tax assessments under Schedule D.
     - (*b*) The *Economist* 'sample of profits' (compiled from the published accounts of a number of companies).

(d) *Method of Construction.*
  (a) Extensive manipulations which are outside the scope of this work are required to adjust the extremely sensitive figures of profits with the very steady figures of income-tax assessments.
  (b) The resulting index number is expressed as:—
    (a) A General Profits Index.
    (b) A Sub-Index for Ordinary Capital.

10. *The* 'Bankers' Magazine' *Investment Index Number.*
  (a) *Use.*—To estimate the level of the prices of stocks and shares.
  (b) *Base Period.*—The quotations as at 31st December, 1921 = 100.
  (c) *Base.*—365 stocks and shares.
  (d) *Method of Construction.*
    (1) The market values of the securities.
    (2) The securities are grouped as follows:—

(a) Fixed interest	87	securities
(b) Dividend	278	,,

    (3) The average prices in each group are equated to 100.
    (4) The base is adjusted as occasion arises, e.g. on liquidation of a company.

11. *The* 'Investors' Chronicle' *Index Number.*
  (a) *Use.*—To estimate the level of the prices of stocks and shares.
  (b) *Base Period.*—The quotations as at 31st December, 1923 = 100.
  (c) *Base.*—176 quotations.
  (d) *Method of Construction.*
    (1) The market values of the securities.
    (2) The securities are grouped as follows:—

Fixed Interest.		
Gilt-edged	19	securities
Business	15	,,
	34	
Variable Dividend.		
First-class business	16	,,
General business	70	,,
Speculative	56	,,
	142	
Total	176	

    (3) Separate index numbers are given for:—
      (a) Fixed interest securities.
      (b) Variable dividend securities.

12. *Wage Index* (London & Cambridge Economic Service: Professor A. L. Bowley).
    - (*a*) *Use.*—The measurement of the average rates of wages.
    - (*b*) *Base Period.*—1914 = 100.
    - (*c*) *Base.*—20 groups of occupations.
    - (*d*) *Method of Construction.*
        - (1) The index is based on the year 1924 equalling 195, when 1914 = 100.
        - (2) Weighting is adopted in respect of selected industries.
        - (3) The index takes into account changes in time and piece rates.
        - (4) The changes in the 20 groups are expressed as percentages of the rates ruling in 1924, the weighted average (an average of ratios) forming the index number.

Chapter IX

DISPERSION AND SKEWNESS

Definition: Use of Graphic Methods: Absolute and Relative Measures: The Range: The Moments of the Dispersion: Average Deviation and Coefficient: Standard Deviation and Coefficient: The Modulus: Quartile Measure and Coefficient: Skewness—1st, 2nd and 3rd Measures and Coefficients: Further Examples.

DISPERSION, alternatively termed 'deviation', refers to the lack of uniformity existing in a mass of data. Different series of data may yield the same average, but the differences of the items from the average may vary considerably. This deviation from a selected type or average is often important from the point of view of statistical research; in fact, the deviations, when making comparisons, are usually of greater importance than the average itself. The ascertainment of dispersion—a measure of how the observations are clustered around a mean—may thus be necessary, either to test the significance of an average, i.e. to what extent the average is typical, or to form the basis of further inquiries.

Dispersion may be illustrated by—

(1) Comparing the deviations of sets of data as tabulated or plotted on a graph.

(2) Making similar comparisons of the data when plotted so as to form a Lorenz curve.

These methods, however, merely indicate the absolute dispersion of the series under consideration, and do not supply a numerical measure of the extent of the variations.

To obtain numerical factors resort has thus to be made to mathematical methods.

Dispersion may be measured absolutely and relatively.

The absolute measure gives a result derived from the actual figures and does not take into account the magnitudes of the items in the series from the point of view of comparability of different masses of data.

The relative measure gives a coefficient or ratio of dispersion, thus facilitating comparison of the variations of widely differing sets of material.

For example, two distributions may give identical measures of dispersion while their means differ. The distribution having the larger mean thus has relatively less variation than the other, when the measure of dispersion (for comparative purposes) should be expressed as a ratio of the mean.

The simplest method of the measurement of dispersion is to ascertain the range of the data, i.e. the difference between the magnitudes of the extreme items. This method is of very little practical use, as it depends on the sizes of extremes (in all probability abnormal items). Similarly, where sampling is employed, it may not give a truly representative measurement.

The moments of the dispersion which form the basis of the measures of dispersion are derived as follows:—

Let d = the deviations of the individual items from the average selected.
n = the number of items.

Then the—

(*a*) 1st moment of the dispersion $= \frac{\Sigma d}{n}$

(*b*) 2nd moment of the dispersion $= \frac{\Sigma d^2}{n}$

(*c*) 3rd moment of the dispersion $= \frac{\Sigma d^3}{n}$

The most usual method of ascertaining the absolute dispersion is to calculate the deviation from the arithmetic average, the mode or the median, whichever basis may be most convenient or useful to the inquiry, although technically the median is the correct type to utilise, the arithmetic average is usually the most convenient, while the mode is rarely used.

The relative measure or coefficient of dispersion may be obtained by dividing the absolute measure of dispersion by the average selected to form the basis of the former calculation.

The measures of dispersion and their coefficients are given below.

(*a*) *The Average Deviation.*

This measure (absolute) which gives weight to extremes, is the first moment of the dispersion, i.e. the arithmetic average of the deviations, and is ascertained by—

(1) Totalling the deviations of the individual items from the average selected, i.e. the arithmetic average, the median, or the mode. (As the algebraic sum of the deviations from the arithmetic average is zero, and from the mode and the median very small, the deviations are all regarded as positive.)

(2) Dividing the sum of the individual deviations by the number of the items.

Formula.

Let δ = the average deviation.
d = the deviations of each individual item from the average.
d_M = the deviations of each individual item from the median.
d_Z = the deviations of each individual item from the mode.
n = the number of items.

Then $\delta = \frac{\Sigma d}{n}$, the average deviation from the arithmetic average.

$\delta_M = \frac{\Sigma d_M}{n}$, the average deviation from the median.

$\delta_Z = \frac{\Sigma d_Z}{n}$, the average deviation from the mode.

The coefficient is ascertained by dividing the absolute measure by the average or type selected.

Formula.

Let a = the arithmetic average.
Then the coefficient of average deviation =

$$\frac{\delta}{a} \quad \text{or} \quad \frac{\delta_M}{M} \quad \text{or} \quad \frac{\delta_Z}{Z}$$

Example 37.

(For the sake of clarity a simple table has been adopted to illustrate the various methods of measuring deviation.)

	Items (1)		Deviations from arithmetic average (2)	Deviations squared (3)
	5		— 11	121
	7		— 9	81
	9		— 7	49
	12		— 4	16
	18		+ 2	4
	21		+ 5	25
	22		+ 6	36
	24		+ 8	64
	26		+ 10	100
	9)144		62	496
Arithmetic average	16	Algebraic sum	0	

Applying the formula.

$$\text{Absolute measure} = \frac{62}{9} = 6.9$$

$$\text{Cöefficient} = \frac{6.9}{16} = .431$$

(*b*) *The Standard Deviation.*

This measure (absolute), which gives additional weight to extremes and is thus greater than the average deviation, is derived from the second moment, viz. $\frac{\Sigma d^2}{n}$, and is ascertained by—

(1) Squaring the individual deviations from the arithmetic average.
(2) Adding the squared numbers.
(3) Dividing the result so obtained (operation (2)) by the number of items.
(4) Extracting the square root of operation (3).

It is thus the square root of the arithmetic average of the squared deviations. Although more difficult to calculate than the average deviation, it is much more valuable than the average or the quartile (*post*) measures, as it admits of further mathematical treatment.

Formula.

Let σ = the standard deviation.

$$\text{Then } \sigma = \sqrt{\frac{\Sigma d^2}{n}}$$

The coefficient is ascertained by dividing the absolute measure by the arithmetic average.

The quantity σ^2 is known as the Variance and is much used in the more advanced forms of statistical analysis.

Formula.

$$\text{The coefficient of standard deviation} = \frac{\sigma}{a}$$

Example 38. Data as in Example 37.

Applying the formula.

$$\text{Absolute measure} = \sqrt{\frac{496}{9}} = 7.4.$$

$$\text{The coefficient} = \frac{7.4}{16} = .463.$$

Alternative Method.—Where the average is not a whole number the work involved in calculating and squaring the deviations is considerable, when the application of the undermentioned method will be found to be far more simple.

(1) Select a whole number of approximate value to the average.
(2) Ascertain the deviations from the assumed average.
(3) Square the numbers so obtained (operation (2)).
(4) Add the squared numbers.
(5) Subtract from (4) n times the square of the difference between the assumed and true average.
(6) Divide the result by n.
(7) Extract the square root.

Formula.

Let x = the assumed average.

Then $$\sigma = \sqrt{\frac{\Sigma d_x^2 - n\,(a - x)^2}{n}}$$

Example 39.

Data as in Example 37, with the exception of the last item, viz :—

	Items (1)	Deviations from Assumed Average (2)	Deviations Squared (3)
	5	— 11	121
	7	— 9	81
	9	— 7	49
	12	— 4	16
	18	+ 2	4
	21	+ 5	25
	22	+ 6	36
	24	+ 8	64
	27	+ 11	121
	9)145	63	517
Arithmetic average	16.11		
Assumed average	16		

Applying the formula.

$$\text{Absolute measure} = \sqrt{\frac{517 - 9(16.11 - 16)^2}{9}}$$

$$= \sqrt{\frac{517 - 9(.11)^2}{9}}$$

$$= \sqrt{\frac{517 - .1089}{9}}$$

$$= 7.5$$

(*c*) *The Coefficient of Variation.*

This quantity is often used when making comparisons of variation and is obtained by expressing the coefficient of standard deviation as a percentage.

Formula.

Let v = the coefficient of variation.

$$\text{Then} \quad v = \frac{\sigma}{a} \times 100$$

Other methods of measuring dispersion are given below:—

(*d*) *The Modulus.*—The modulus, which is based on the second moment and which gives weight to extremes, is ascertained by—

(1) Squaring the individual deviations from the arithmetic average.
(2) Adding the squared numbers.
(3) Multiplying the result of (2) by 2.
(4) Dividing the result of (3) by the number of items.
(5) Extracting the square root of operation (4).

This measure, however, is not in common use.

Formula.

$$\text{The modulus} = \sqrt{\frac{2\ \Sigma d^2}{n}}$$

Example 40. (Data as in Example 37.)

Applying the formula.

$$\text{The modulus} = \sqrt{\frac{992}{9}} = 10.5.$$

(*e*) *Quartile Measure of Dispersion.*—This measure, which gives no weight to extremes and is thus less than the average and standard measures, is ascertained by—

(1) Subtracting the magnitude of the 1st quartile from that of the 3rd quartile.
(2) Dividing the result of operation (1) by 2.

The difference between the quartiles is known as the quartile range, and half the quartile range is termed the semi-interquartile range or quartile deviation.

Formula.

Let Q_1 = the 1st quartile.
Q_3 = the 3rd quartile.

Then the quartile measure of dispersion $= \dfrac{Q_3 - Q_1}{2}$.

The usual method of obtaining the coefficient is by dividing the quartile deviation by half the sum of the 1st and 3rd quartiles.

Formula.

$$\text{The coefficient} = \frac{\dfrac{Q_3 - Q_1}{2}}{\dfrac{Q_3 + Q_1}{2}} = \frac{Q_3 - Q_1}{Q_3 + Q_1}.$$

Example 41. (Data as in Example 37.)

Data arrayed.

5
7 1st quartile = 8
—
9
12
18 Median = 18
—
21
22 3rd quartile = 23
—
24
26

Applying the formula.

The quartile measure of dispersion $= \dfrac{23 - 8}{2} = 7.5$.

The coefficient $= \dfrac{15}{23 + 8} = .484$.

(*f*) *Measurement of Skewness.*—Where the data is evenly distributed the magnitude of the arithmetic average, the mode and the median will be the same. (See Chapter XI). Where skewness is present the mode remains at the position of greatest density, the median varies in proportion to the number of irregular items, and the arithmetic average to a greater extent, proportional to the magnitude of the irregular items.

A vertical line drawn from the x axis of a distribution from—
(a) the mode, cuts the position of greatest ordinate;
(b) the median, bisects the area;
(c) the arithmetic average, passes through the centre of gravity of the histogram. Examples, Figures 29 and 30.

The measurement of skewness, i.e. the absence of symmetry, as distinct from dispersion—the deviations from a type—appearing in frequency curves is effected by three methods, viz.:—

(1) *The 1st Measure of Skewness*, which takes the whole of the data into account. This measure is ascertained by deducting the magnitude of the mode from the arithmetic average.

Note.—(1) Where the mode is badly defined the median may be substituted, although the latter is not such a satisfactory basis as the mode.

(2) Where the median is utilised the resulting measure is smaller than that resulting from the use of the mode.

Formula. Let Z = the mode.
M = the median.
a = the arithmetic average.
Then the 1st measure of skewness = a — Z or a — M.

Where the arithmetic average travels to the right, as illustrated in a curve, there is positive, and to the left negative, skewness. Figures 29 and 30.

The coefficient is ascertained by dividing the absolute measure by the mean or alternatively the standard measure of deviation. The difference in principle in this calculation when compared with that regarding dispersion, lies in the fact that in the case of dispersion the measurement is based on deviations from an average, while in the case of skewness it is desired to measure the difference between the deviations arising on both sides of the average.

Formula.

Let j = the coefficient of skewness.

$$\text{Then } j_1 = \frac{a - Z}{\delta} \text{ or } \frac{a - Z}{\sigma}.$$

Note.—Where the mode is utilised in the calculation to ascertain the measure of skewness, it must also be used in calculating the coefficient, similarly in the case of the median.

It has been observed that, in the case of many moderately asymmetrical distributions, the median travels to a position about two-thirds the distance the arithmetic average is displaced. In such cases the undermentioned formula may be suitable where the mode is badly defined.

Formula.

$$j_1 = \frac{3(a - M)}{\delta} \text{ or } \frac{3(a - M)}{\sigma}.$$

Example 42.

Data as in Example 37 (the median is utilised in the calculation). Applying the formula.

The 1st measure of skewness $16 - 18 = -2$.

$$\text{The coefficient} = \frac{-2}{6.9} = -.029.$$

(2) *The 2nd Measure of Skewness.*—This measure, which ignores extreme items, falling outside the positions of the two quartiles and is thus smaller than the 1st measure, is ascertained by adding the magnitudes of the 1st and 3rd quartiles and deducting therefrom a sum equal to twice the size of the median.

It thus measures the position of the quartiles in relation to the median, which, in the presence of skewness does not lie exactly half-way between the quartiles.

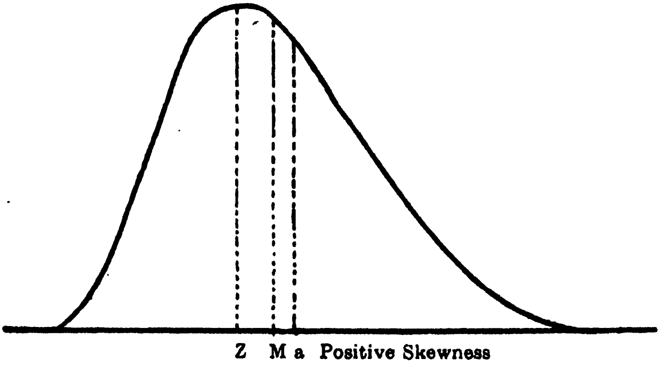
Formula.

The 2nd measure of skewness $= Q_3 + Q_1 - 2M$.
The method of obtaining the coefficient is by dividing the absolute measure by the quartile measure of deviation.

Figure 29.

DIAGRAM SHOWING SKEWNESS

(Moderately Asymmetrical Distribution)



Formula.

$$j_2 = \frac{Q_3 + Q_1 - 2M}{\frac{Q_3 - Q_1}{2}}$$

Example 43.

Data as in Example 37.

$$\text{The 2nd measure of skewness} = 23 + 8 - 36$$
$$= -5$$

$$\text{The coefficient} = \frac{-5}{7.5} = -.061.$$

(3) *The 3rd Measure of Skewness.*—This measure is based on the 3rd moment of the distribution and may be expressed as follows:—

Formula.

The 3rd measure of skewness $\sqrt[3]{\frac{\Sigma d^3}{n}}$.

The third measure gives emphasis to the extreme items and is greater than the 1st and 2nd measures.

The method of obtaining the coefficient is by dividing the absolute measure by the standard deviation.

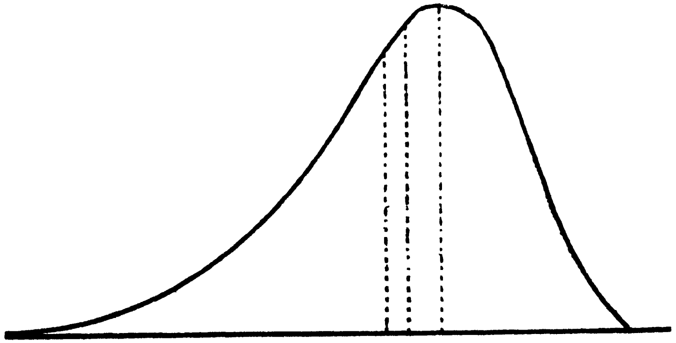
Formula.

$$j_3 = \frac{\sqrt[3]{\frac{\Sigma d^3}{n}}}{\sigma}$$

Figure 30.

DIAGRAM SHOWING SKEWNESS

(Moderately Asymmetrical Distribution)



Negative Skewness

Example 44. Data as in Example 37.

The 3rd measure of skewness = 7.8.

The coefficient $= \frac{7.8}{7.4} = 1.054.$

DISPERSION

A further example is given of the calculation of measures of dispersion and skewness, in this case from a frequency distribution.

Example 45. Original data as in Example 27 (averages).

Item	Value *	Mid-Value (m)	$f.$	$d.$	$fd.$	($x = 10$) d_x	d_x^2	fd_x^2
1	2–4	3	1	7.24	7.24	− 7	49	49
2	4–6	5	6	5.24	31.44	− 5	25	150
3	6–8	7	15	3.24	48.60	− 3	9	135
4	8–10	9	20	1.24	24.80	− 1	1	20
5	10–12	11	15	0.76	11.40	+ 1	1	15
6	12–14	13	10	2.76	27.60	+ 3	9	90
7	14–16	15	6	4.76	28.56	+ 5	25	150
8	16–18	17	4	6.76	27.04	+ 7	49	196
9	18–20	19	2	8.76	17.52	+ 9	81	162
			79		224.20		249	967

* i.e., over 2 and not exceeding 4, &c.

Note.—$a = 10.24$ (Example 27).

The Average Deviation $= \frac{\Sigma d}{n} = \frac{224.20}{79}.$

$\delta = 2.85.$

The coefficient of average deviation $= \frac{\delta}{a} = \frac{2.85}{10.24}.$

$\frac{\delta}{a} = 0.28.$

The Standard Deviation $= \sqrt{\frac{\Sigma d^2}{n}} = \sqrt{\frac{\Sigma d_x^2 - n(a - x)^2}{n}}$ (as in this example the deviations are taken from the assumed average of 10) $= \sqrt{\frac{967 - 79(10.24 - 10)^2}{79}}$

$\sigma = 3.48.$

The coefficient of standard deviation $= \frac{\sigma}{a} = \frac{3.48}{10.24}$

$\frac{\sigma}{a} = 0.34.$

The Quartile Measure of Deviation $= \dfrac{Q_3 - Q_1}{2} = \dfrac{12.60 - 7.73}{2}$

Note.—$Q_3 = 12.60$ (Example 27)
$Q_1 = 7.73$

Q.D. $= 2.44$.

The coefficient of quartile deviation $= \dfrac{\dfrac{Q_3 - Q_1}{2}}{\dfrac{Q_3 + Q_1}{2}} = \dfrac{4.87}{20.33}$

The coefficient of quartile deviation $= 0.24$.

The 1*st Measure of Skewness* $= a - Z = 10.24 - 9$.

The 1st measure of skewness $= 1.24$.

The coefficient of skewness $= \dfrac{a - Z}{\delta} = \dfrac{1.24}{2.85}$

$j_1 = 0.44$

The 2*nd Measure of Skewness* $= Q_3 + Q_1 - 2m = 12.60 + 7.732 - (9.80)$.

The 2nd measure of skewness $= 0.73$.

The coefficient of skewness $= \dfrac{Q_3 + Q_1 - 2m}{\dfrac{Q_3 - Q_1}{2}} = \dfrac{0.73}{2.44}$

$j_2 = 0.29$.

Chapter X

CORRELATION

Definition: Nature: Use of Graphic Methods: Coefficient of Correlation: Regression—Ratio of Variation: General Trend: Short Term Fluctuations: Time Lag.

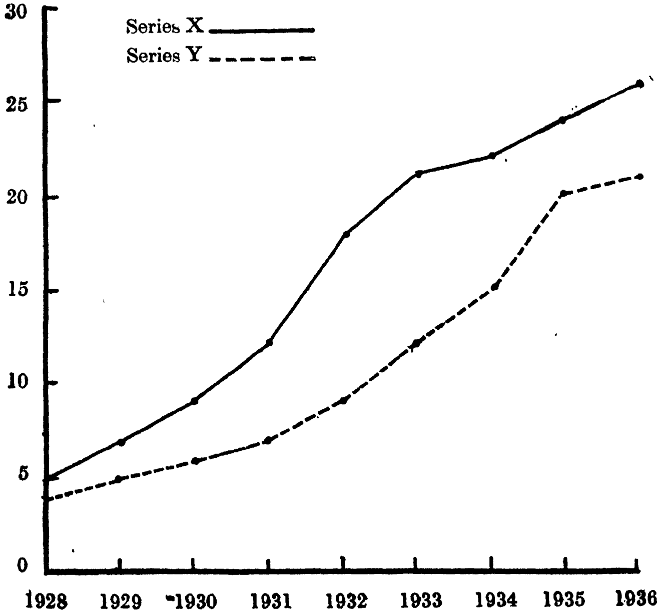
Correlation is the term used to express the fact that a causal connection exists between groups of data, i.e. that a certain 'cause', movement of tendency arising in one direction produces proportionate or sympathetic effects in another. A tendency in one set of data is thus accompanied in greater or less degree by a similar or inverse tendency in the other.

The results of the application of one or more of the tests set out hereafter merely give numerical or pictorial indications of correlation. It must, therefore, be proved that—

(1) The movements in one set of data are directly caused by movements in the other set, i.e. that the factors are definitely related,

Figure 31.

Simple Graph Showing Evidence of Correlation



e.g. the bank rate in relation to the bank reserve, or

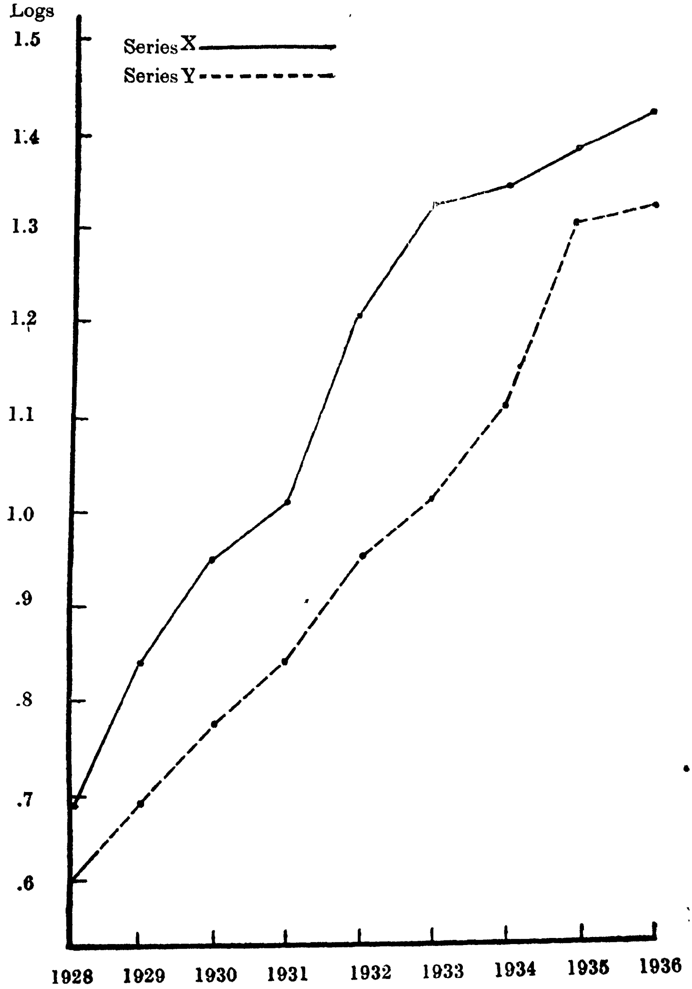
(2) that both sets of data are affected by another factor which is related to both, e.g. the effect of the amount of currency in circulation on the prices of various commodities.

If the variables move in the same direction, the correlation between them is termed 'direct'; if in contrary direction, 'inverse'.

For many practical purposes correlation may be investigated by:—

Figure 32.

LOGARITHMIC GRAPH SHOWING EVIDENCE OF CORRELATION



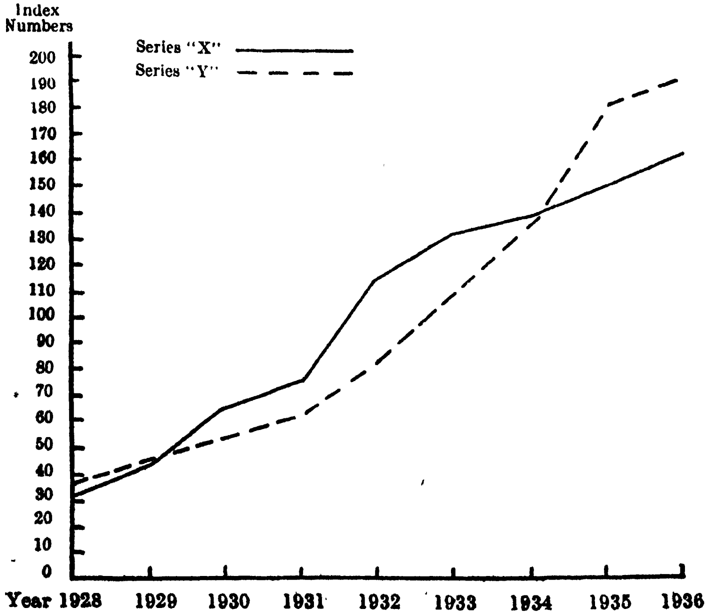
(1) Plotting the factors as a simple graph and noting the general trend of the resulting curves. Example, Figure 31.

DATA

Year	Items 'X'	Logs	Indices	Items 'Y'	Logs	Indices
1928 ..	5	.6990	31	4	.6021	36
1929 ..	7	.8451	44	5	.6990	45
1930 ..	9	.9542	56	6	.7782	55
1931 ..	12	1.0792	75	7	.8451	64
1932 ..	18	1.2553	113	9	.9542	82
1933 ..	21	1.3222	131	12	1.0792	109
1934 ..	22	1.3424	138	15	1.1761	136
1935 ..	24	1.3802	150	20	1.3010	182
1936 ..	26	1.4150	162	21	1.3222	191
	144		9)900	99		9)900
			100			100
Arithmetic average	16			11		

Figure 33.

GRAPH OF INDEX NUMBERS SHOWING EVIDENCE OF CORRELATION



Note.—The arithmetic average has been taken as 100, the base of the index.

(2) Similarly plotting the factors, but as a semi-logarithmic graph. Example, Figure 32.

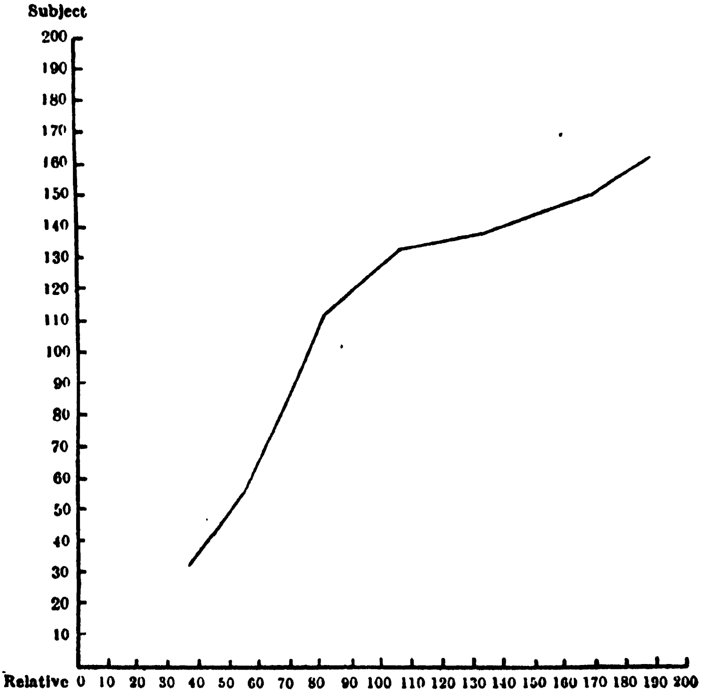
(3) Reducing the series to indices and plotting the indices on a graph (two curves). The trend of the data may then be observed, and if necessary, the fluctuations from the trend or moving average may be similarly plotted and the resulting curves compared. Example, Figure 33.

(4) Constructing a diagram in the form of the Galton graph, by plotting the indices, utilising the vertical scale for one set of data, Series 'X' (the subject) and the horizontal scale for the other set Series 'Y' (the relative).

Figure 34.

DIAGRAM IN THE FORM OF THE GALTON GRAPH

SHOWING EVIDENCE OF CORRELATION



The line formed by joining the points may be interpreted as follows:—

(*a*) A straight line—perfect correlation.

(*b*) A regular curve—perfect correlation, one set of data being subject to a time lag.

(*c*) Should the line drop towards the left-hand side of the diagram—direct correlation.

(*d*) Should the line drop towards the right-hand side of the diagram—inverse correlation.

(*e*) Should an irregular line result, possessing no definite characteristics—no evidence of correlation.

Example, Figure 34.

These graphs may be interpreted as showing evidence of direct correlation between the two sets of data.

The foregoing graphic methods afford some indication of correlation between the variables. It is, however, desirable to obtain some mathematical measure of that relationship. It has been found that investigations of this nature resolve themselves into two problems:—

(1) The amount of dependence arising between the variables.

(2) The probable movements of one variable in terms of the other.

These relationships respectively may be considered by an investigation of—

(1) Correlation.

(2) Regression.

The consideration of correlation is appropriate when the variables are known to be associated, but where one is not the consequence of the other, i.e. both are associated through the operation of a combination of causes.

The consideration of regression, on the other hand, is appropriate when one variable is dependent on the other, i.e. where values of one are wholly or partly caused or controlled by the other.

The relationships of variables may thus be investigated by methods applicable to the consideration of correlation or regression or both.

1. *Coefficient of Correlation.*

A mathematical formula, known as the Karl Pearson coefficient of correlation, has been devised to accurately measure the correlation existing between variables, when the coefficient or correlation represents the amount of the relationship between the sets of variables.

In this formula one series of data (the subject — X) is utilised as the basis of the comparison with which the other series (the relative — Y) is arranged.

The application of the formula involves the following operations.

(1) Array the subject and relative series.

(2) Tabulate the deviations of each item in the subject from its arithmetic average.

(3) Tabulate the deviations of each item in the relative from its arithmetic average.

(4) Multiply each of the deviations in the subject by the corresponding deviations in the relative.

(5) Add the products.

(6) Ascertain the standard deviation of the subject and relative.

(7) Multiply the standard deviation of the subject by the standard deviation of the relative.

(8) Ascertain the number of items under comparison.

(9) Multiply the product obtained in operation 7 by the number of items, operation 8.

(10) The coefficient is obtained by dividing the results of operation 5 by the result of operation 9.

Formula.

Let r = Coefficient of correlation.
x = Deviations of subject items from their arithmetic average.
y = Deviations of relative items from their arithmetic average.
n = The number of pairs of items.
σ_x = The standard deviation of the subject.
σ_y = The standard deviation of the relative.

$$\text{Then } r = \frac{\Sigma\ (xy)}{n\ \sigma_x\ \sigma_y}.$$

The probable error in the coefficient of correlation, arising from errors of sampling, may be obtained by:—

(1) Deducting the square of the coefficient of correlation from unity.

(2) Multiply the result of (1) by .6745 (a mathematical constant—see Chapter XI).

(3) Divide the result of operation (2) by the square root of the number of items.

Formula.
$$\frac{.6745\ (1 - r^2)}{\sqrt{n}}$$

The coefficient so obtained may be interpreted as follows:—

If $r = 0$ no correlation exists.
,, $r = 1$ complete direct correlation.
,, $r = -1$ complete inverse correlation.
,, r is above .5—decided evidence of correlation.
,, r is less than .3—little evidence of correlation.
,, r is less than the probable error—no evidence of correlation.
,, r is more than 3 times the probable error—evidence of correlation.
,, r is more than 6 times probable error—decided evidence of correlation.

Example 46.

DATA

	Subject			*Relative*		
Items 'X'	Deviations from Arithmetic Average 'x'	Deviations Squared 'x^2'	Items 'Y'	Deviations from Arithmetic Average 'y'	Deviations Squared 'y^2'	'xy'
5	— 11	121	4	— 7	49	77
7	— 9	81	5	— 6	36	54
9	— 7	49	6	— 5	25	35
12	— 4	16	7	— 4	16	16
18	+ 2	4	9	— 2	4	— 4
21	+ 5	25	12	+ 1	1	5
22	+ 6	36	15	+ 4	16	24
24	+ 8	64	20	+ 9	81	72
26	+ 10	100	21	+ 10	100	100
144	62	496	99	48	328	383
						— 4
						379

Arithmetic average ..	16	11
Standard deviation ..	7.42	6.04
No. of items	9	9

Applying the formula.

$$r = \frac{379}{9 \times 7.42 \times 6.04} = \frac{379}{403.35} = .939.$$

The probable error in the coefficient is:—

$$\text{Applying the formula} = \frac{.6745\ (1 - .939^2)}{\sqrt{9}}$$

$$= \frac{.6745\ (1 - .8817)}{3} = \frac{.6745 \times .1183}{3}$$

$$= .0266.$$

The coefficient of correlation may thus be stated as .939 ± .027, i.e. decided evidence of direct correlation.

2. *Regression Line.*

The relationship of the factors X and Y may be exhibited by forming a diagram in which the points are plotted by reference to the vertical axis for the subject and the horizontal axis for the relative. Figure 35 (crosses).

Alternatively, instead of using the actual numbers, index numbers may be plotted.

A 'line of best fit' is then drawn through and as near to all the observations as possible. This line of best fit or regression line is obtained by transferring the origins of X and Y to their means.

$$\text{Then } x = \mathrm{X} - a_x$$
$$y = \mathrm{Y} - a_y$$

The regression line is thus the line which indicates the trend of the combined observations.

The regression equation expresses the most probable value of y associated with a given value of x, and is devised to make, the squares of the vertical distances from the points to the line and the original observations, a minimum.

Formula. Regression equation.

$$\mathrm{Y} - a_y = r \frac{\sigma_y}{\sigma_x} (\mathrm{X} - a_x)$$

i.e. the regression line of y on x or the 'line of best fit' of Y to X.

Example 47. Regression Equation.

Data as in Example 46.

$$\mathrm{Y} - 11 = .939 \times \frac{6.04}{7.42} (\mathrm{X} - 16)$$
$$= y = .939 \times \frac{6.04}{7.42} (x)$$
$$= y = .765x - 1.24.$$

Then, substituting values of x in the equation:

When	$x = 5$	18	26
	$y = 2.6$	12.5	18.6

These values are plotted in Figure 35 where there sulting regression line is indicated by a continuous line.

If the correlation is perfect, the subject and relative vary in unison and where correlation is not perfect there is a tendency for variation in unison.

This measure, however, does not give an indication of the relative extent of the movements, i.e. whether variations in the subject are associated with an equal, smaller or greater variation in the relative.

These relative movements are investigated by obtaining an average ratio between the proportional deviations of the subject and the relative.

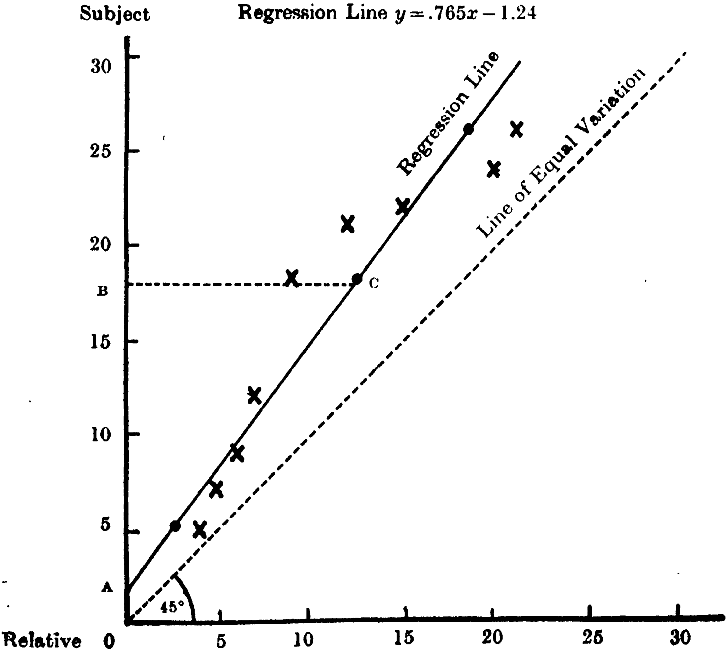
This factor is known as the ratio of variation.

Ratio of Variation.

The slope of the regression line gives the average ratio of variation of the variables.

1. Where the subject and relative change by an equal ratio.	The line will be at an angle of 45° to the horizontal, when the ratio of variation is equal to 1.
2. Where the subject changes more in proportion than the relative.	The line will be at an angle greater than 45° to the horizontal.
3. Where the subject changes less in proportion than the relative.	The line will be at an angle less than 45° to the horizontal.

Figure 35.



GALTON GRAPH

A line of 45° to the horizontal is termed the line of equal variation. Figure 35 (broken line).

The numerical value of the ratio of variation is the tangent of the angle that the regression line makes with the vertical.

This ratio may be obtained by dividing the horizontal distance of any point B from this line by its vertical distance,

i.e. (see Figure 35) $\frac{BC}{BA} = \frac{12.5}{16.5} = .757.$

This means that for every change of 1 in the subject, there is a tendency for the relative to change by .757.

The whole diagram, Figure 35, is termed a Galton graph.

In the example given the correlation of the general trend of the data has been investigated, and consequently deviations from the arithmetic average have formed the basis of the calculation.

It is possible, however, that the trends of two series may be correlated directly while their short term fluctuations are correlated inversely, or vice versa.

Where it is desired to consider short term fluctuations the deviations should be calculated from the trend. This is achieved for the subject and relative by:—

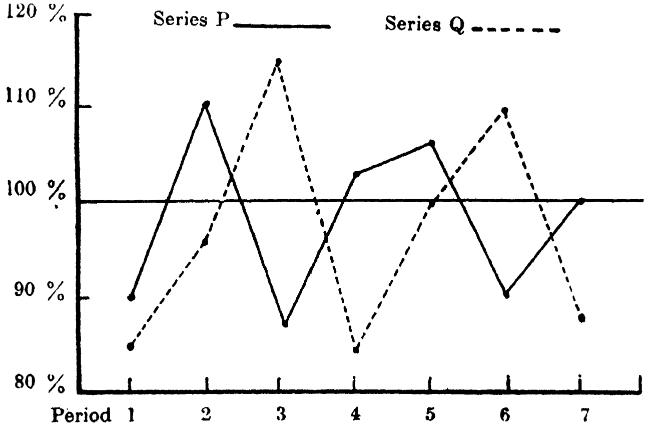
(*a*) calculating the deviations from a moving average, or

(*b*) (1) Reducing the factors to index numbers.
  (2) Calculating a moving average of the index numbers.
  (3) Utilising the deviations of the index numbers from the moving average as the basis for calculating the standard deviations.

In other respects the formula remains unchanged.

Figure 36.

GRAPH SHOWING EVIDENCE OF CORRELATION WITH TIME LAG



Time Lag.—The time factor between the cause of the change and when it was effected may arise when investigating correlation, when it would be necessary to allow for the 'time lag', i.e. the period elapsing between 'a cause' and its 'effect', in tabulating the facts under review.

Figure 36 illustrates a case where correlation is indicated but there is a time lag between the curves.

In such cases it is necessary to establish which causes the effect in the other.

For example:—

P does not necessarily cause Q
Q may cause P or
Q may cause minus P.

DATA

Example 48.

Period	Series P	Series Q
1	90	85
2	110	96
3	87	115
4	103	85
5	106	100
6	90	110
7	100	88

Chapter XI

THE NORMAL FREQUENCY CURVE

Probability: Introduction: Coin Tossing: Binomial Expansion: Normal Frequency Curve: General Description: Characteristics: Approximate Relationship of Measures of Deviation: Sampling: Law of Statistical Regularity: Law of Inertia of Large Numbers: Statistical Significance: Non-Measurable Characteristics: Measurable Characteristics: Distribution of the Means of Samples: Significance of the Difference between Means of Samples.

In this chapter a brief description is given of the normal frequency distribution and some of its implications.

1. *Probability.* Probability in the statistical sense implies a measure of expectation that an event will or will not occur. It has been observed that a certain amount of regularity occurs in the proportionate frequency of the occurrence of events, when the circumstances under which they occur are repeated many times, i.e. that when repeated trials are made under similar conditions, the proportionate frequency of the occurrence of a particular event appears to be constant.

From the foregoing it has been inferred that:—

Definite occurrences are associated with certain circumstances, or

Each circumstance gives rise to either the occurrence or the non-occurrence of a definite event, or

A set of circumstances give rise to conditions favourable to or adverse to the occurrence of an event.

The foregoing suggest the possibility of arguing back from the measured frequency of an event (e.g. the number of deaths in a community), to the nature of the system in which these occurrences arise (e.g. is the community different from others or from itself over a period of time?).

In discussing probability in practical forms it is necessary to make working assumptions.

Consider a six-sided die:—

(1) In speaking of the probability of throwing a 'six', the statistician is not concerned with metaphysical conceptions of the absolute chance of this event occurring at the next trial, but with the relative frequency of this event occurring in a large number of throws.

(2) It is assumed that if a certain set of circumstances gives rise to a 'six', then, if those circumstances are exactly repeated, a 'six' will always occur.

Suppose that an event can occur in n ways, s of which are considered successes, and f as failures, so that $s + f = n$. Then if it can be assumed that all ways are equally likely it is possible to calculate the relative frequency or probability (p) of a success by the equation

$$p = \frac{s}{n}$$

and the probability of a failure (q), by

$$q = \frac{f}{n}$$

Then $p + q = 1$—a certainty that success or failure must always occur, and $s + f = n$.

To use familiar examples:—

(1) The probability of obtaining a head on tossing a coin is 1 in 2 ($p = \frac{1}{2}$, $q = \frac{1}{2}$).

(2) The probability of obtaining an ace from a pack of playing cards is 4 in 52 ($\frac{1}{13}$) in favour, or 48 in 52 ($\frac{12}{13}$) against this event happening. ($p = \frac{4}{52}$, $q = \frac{48}{52}$).

To extend the example of coin tossing:—

If 2 coins are tossed there are 2 events, each of which may occur in 2 ways, i.e. 4 possible results ($2^2 = 4$).

	Chance	Ratio
H.H.	$\frac{1}{4}$	1
H.T., T.H.	$\frac{1}{2}$	2
T.T.	$\frac{1}{4}$	1
	1	

Similarly with 3 coins there are 3 events, each of which may occur in 2 ways, i.e. 8 possible results ($2^3 = 8$).

	Chance	Ratio
H.H.H.	$\frac{1}{8}$	1
H.H.T., T.H.H., H.T.H.	$\frac{3}{8}$	3
H.T.T., T.H.T., T.T.H.	$\frac{3}{8}$	3
T.T.T.	$\frac{1}{8}$	1
	1	

With 4 coins there are 4 events, each of which may occur in 2 ways, i.e. 16 possible results ($2^4 = 16$).

	Chance	Ratio
H.H.H.H.	$\frac{1}{16}$	1
H.H.H.T., H.H.T.H., H.T.H.H., T.H.H.H.	$\frac{4}{16}$	4
H.H.T.T., H.T.H.T., T.H.H.T., H.T.T.H., T.H.T.H., T.T.H.H.	$\frac{6}{16}$	6
H.T.T.T., T.H.T.T., T.T.H.T., T.T.T.H.	$\frac{4}{16}$	4
T.T.T.T.	$\frac{1}{16}$	1
	1	

Each successive series of ratios thus progress in a uniform manner, as shown by the following table:—

1 event in 2 ways	1, 1
2 events in 2 ways	1, 2, 1
3 events in 2 ways	1, 3, 3, 1
4 events in 2 ways	1, 4, 6, 4, 1

It will be noticed that the ratios shown in the above examples are the coefficients in the expansion of $(p + q)^n$, where n is the number of coincident events considered.

The expansion, for example, of $(p + q)^4$ gives rise to :—

$$p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$$

In each case the coefficient is associated with the combination of probabilities actually involved. Thus the chance of H.H.T.T. is $p \times p \times q \times q$, i.e. p^2q^2, and this has the coefficient 6 in the above expansion, i.e. its relative frequency compared with the frequency occurrence of (for example), H.H.H.H. (p^4), taken as 1.

This general method of calculating the frequency distribution of successes and failures is important.

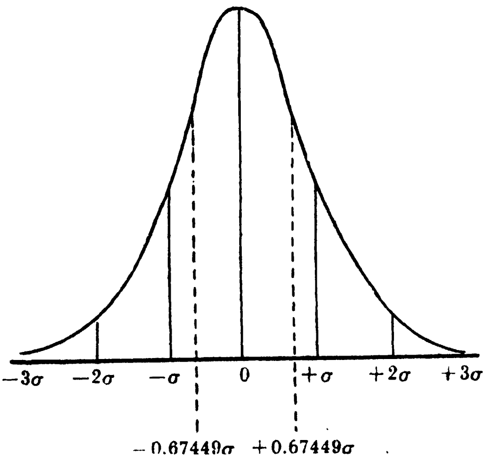
A distribution thus calculated is known as the binomial distribution.

The above principles, moreover, may be applied to any combination of events—it is immaterial whether $p = q$, but only that $p + q = 1$.

These ratios when plotted as frequency distributions, form symmetrical bell-shaped curves. This form results from the fact that the position of each item in the curve is determined by several factors.

Figure 37.

THE NORMAL FREQUENCY CURVE



These factors may be

(1) all present;
(2) all absent;
(3) present or absent in greater or less degree (intermediate positions).

On the average they tend to compensate each other, where they do not compensate the items resulting occur at the extremes of the curve.

2. *The Normal Frequency Curve.*

When the number of items under consideration (n) is large and p and q approximately equal, the expansion of the term $(p + q)^n$ by the binomial theorem, approximates to a curve known as the normal frequency curve, alternatively termed the normal curve of error or the normal probability curve. This curve, which can be defined by a mathematical formula, is symmetrical, continuous and bell-shaped. It rises to a high peak at the position of greatest ordinate and falls rapidly towards the axis of x, whence it stretches (mathematically) to infinity in both directions.

An illustration of the characteristic bell-shape of the normal frequency curve is given in Figure 37.

When plotted as an ogive, the curve is S-shaped, Figure 38.

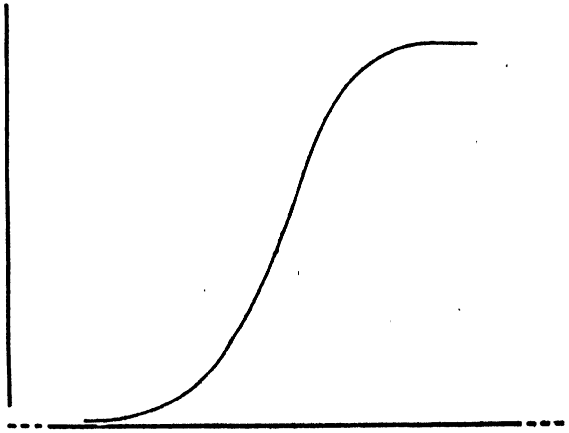
Being symmetrical the arithmetic average, the mode and the median coincide and the question of skewness does not arise.

The mean of the distribution is pn, and the standard deviation $\sqrt{pqn}$.

The quartiles, which include half the number of observations, are situated at points 0.67449 times the standard deviation, on either side of the mean.

Figure 38.

THE NORMAL FREQUENCY OGIVE



Other calculated factors in relation to the standard deviation are given below:—

Deviation	Proportion of total items included	Percentage of total area of curve included
$+\ 0.5\sigma$ and $-\ 0.5\sigma$	0.383	38%
$+\ 1.0\sigma$ and $-\ 1.0\sigma$	0.682	68%
$+\ 1.5\sigma$ and $-\ 1.5\sigma$	0.866	87%
$+\ 2.0\sigma$ and $-\ 2.0\sigma$	0.954	95%
$+\ 2.5\sigma$ and $-\ 2.5\sigma$	0.988	99%
$+\ 3.0\sigma$ and $-\ 3.0\sigma$	0.997	99.7%
$+\ 0.67449\sigma$ and $-\ 0.67449\sigma$	0.500	50%

The foregoing indicates that, in a normal frequency distribution, there is, for example, a chance of 682 in 1,000 that an item will not exceed the standard deviation and 318 in 1,000 that it will. The chances of an item falling within or outside the area bounded by the quartiles are equal, hence the importance of the quantity 0.67449σ, which is known as the probable error of the distribution, i.e. deviation above or below which actual deviations are equally likely.

For many purposes the observations may be regarded as being contained in an area bounded by plus and minus three times the standard deviation, the probability of an observation exceeding these limits being only 3 in 1,000. Thus the standard deviation should be approximately one-sixth of the range of the data.

Other approximate relationships are given below:—

The average deviation approximates to $\frac{4}{5}$ths of the standard deviation.

The quartile deviation approximates to $\frac{2}{3}$rds of the standard deviation.

99 % or more of the observations lie within a range of:—

6 times the standard deviation.
7.5 times the average deviation.
9 times the quartile deviation.

The probability of a deviation greater than:—

σ is 1 in 3 items approximately
2σ is 1 in 20 items approximately.
3σ is 1 in 370 items approximately.

If *a priori* reasons are held for assuming the data under consideration to be associated with such a distribution, a considerable amount of valuable information as to range and dispersion is readily available, without the process of making a large number of observations. This is the basis of sampling methods.

3. Attention is now drawn to two practical facts, which are generally referred to as 'laws', viz.:—

1. The Law of Statistical Regularity.
2. The Law of the Inertia of Large Numbers.

The former states that if a large sample is taken at random from data it will be representative of the characteristics of the entire population (universe), and the latter that in large masses of data abnormalities will occur, but in all probability, exceptional items will offset each other, leaving the average unchanged (subject, where the element of time enters, to the general trend of the data).

If each and every item of a community has an equal chance of inclusion in the sample, it can be shown from what has already been stated regarding probabilities, that:—

1. A sample will give the same frequency distribution as the population, i.e. it will be truly representative.
2. Any one sample may diverge from the population. The divergence itself, in respect of any characteristic, follows the law of the normal curve, and can be calculated.

The larger the sample, the smaller is the relative range of fluctuation. The reliability of a sample is proportional to the square root of the number of items it contains, and the larger the sample the more representative and stable it will be.

Natural, scientific and to a large extent, economic data follow the calculated normal—a bell-shaped curve results, although most natural and other curves tend to slight skewness in one direction or the other, and, unless a sufficiently representative number of items is taken, to irregularities in shape.

A comparison of the factors relating to coin tossing and those relating to other phenomena is given below:—

Coins.	Other phenomena.
1. The number of classes is fixed (i.e. the number of coins, plus one).	1. Classes are chosen by personal selection.
2. The effect of each factor is equal.	2. Factors may not be equal. Each in fact, may vary—on a bell-shaped curve with or without the presence of skewness—or overshadow other factors.
3. Factors are independent of each other.	3. Factors may not be independent.
4. The chances of operation for each factor are equal.	4. The chances for each factor may not be equal.

Subject, however, to the fact that very large numbers may give the characteristics of the coin experiment, owing to compensating factors arising to offset each other.

Although all the conditions precedent to the normal frequency curve are rarely found in practice, it provides a useful basis upon which to investigate and analyse physical and economic phenomena. Thus if the probabilities are known it is possible to calculate the probable frequencies of the number of successes or failures and to compare these calculated factors with the factors obtained from the investigation of the characteristics shown by samples of the actual data.

Before considering the application of probability to test the significance of results and to assess the limits of error, it is necessary to consider the question of statistical significance.

4. *Statistical Significance.*

The term 'statistically significant' refers to the fact that if differences between the mean and other measures, derived from samples presumed drawn from the same population, exceed certain limits, those differences cannot be attributed to chance as a result of random sampling and a specific cause must be sought by further investigation.

Thus those differences are significant of one of two alternatives:—

(1) They arise as the result of random sampling.

(2) They arise by reason of the fact that the samples have not been drawn from the same population.

Where differences are large it may be highly improbable that the samples have been drawn from the same population, in which case the sets of data are not homogeneous and therefore not comparable.

The foundation of tests of statistical significance rests on the basis of the characteristics of the normal frequency curve.

The chance of 19 to 1 against an occurrence resulting from random sampling, or, in other words—bearing the characteristics of the normal curve in mind—twice the standard error (deviation) or three times the probable error (quartile deviation), has been considered justifiable for the limits in considering statistical significance. It is therefore reasonable to assume that a difference between an expected (calculated) value and the actual value is statistically significant if it exceeds a sum equal to twice the standard deviation or three times the quartile deviation of the distribution under consideration. Where such a difference occurs further classification and analysis should be undertaken, as in all probability the items comprising the data, although possessing some similar characteristics, have other features which do not render the aggregate wholly homogeneous.

Where the population is subject to non-measurable characteristics the object in investigating the sample is to find the proportion possessing the attribute and to apply it to the population. In the case of measurable characteristics (variables) the object is to estimate the mean and measures of dispersion and skewness.

5. *Non-measurable Characteristics (Attributes).*

It is desired to find the expectation of the number of successes in the population when the expression $p = \frac{s}{n}$ represents the proportion of successes in the sample.

The chances of 0, 1, 2 . . . . $n - 2$, $n - 1$, n, successes are given by the terms of the expansion of $(p + q)^n$. The expression $(p + q)^n$ can be represented by a frequency distribution with:—

$$\text{Mean} = pn$$

$$\text{Standard deviation} = \sqrt{pqn}$$

The measure of expectation of the number of successes in a sample of n, where $\sqrt{pqn}$ represents the standard error or measure of unreliability of pn, is

$$pn \pm \sqrt{pqn}$$

The measure of expectation of the proportion of successes is

$$p \pm \sqrt{\frac{pq}{n}}$$

The best estimate, therefore, of the proportion of successes in the whole population (N) is given by the expression:—

$$Np \pm \sqrt{\frac{pq}{n}} \times N$$

The foregoing expressions have been given on the basis of the standard error, when working from the probable error the estimate of the proportion of successes in the population is—

$$Np \pm 0.67449 \sqrt{\frac{pq}{n}} \times N$$

For example, out of a sample of 910 observations, 273 possess a certain characteristic. It is desired to find the expected number possessing that characteristic in a population of 10,000.

Example 49.

$$n = 910$$

$$pn = 273$$

$$p = .3 \text{ i.e. } \left(\frac{273}{910}\right)$$

$$q = .7 \text{ i.e. } \left(\frac{637}{910}\right)$$

$$\sqrt{\frac{pq}{n}} = .0152$$

Then the proportion of the population possessing the characteristic

$$= p \pm \sqrt{\frac{pq}{n}} = .3 \pm .0152$$

And the expected number in the population possessing the characteristic $= Np \pm \sqrt{\frac{pq}{n}} \times N$

$$= 10{,}000 \times .3 \pm .0152 \times 10{,}000$$
$$= 3{,}000 \pm 152$$

The chance of the estimated proportion differing from the actual by $\sqrt{\frac{pq}{n}}$ (.0152 in the example), is 318 in 1,000, or .3 approximately.

When working on the basis of the probable error the chance of the estimated proportion differing from the actual by $.6745\sqrt{\frac{pq}{n}}$ (.0103 in the example) is .5.

Applying the limits previously indicated, i.e. twice the standard or three times the probable error, the expected number in the population possessing the characteristic is $3{,}000 \pm 304$, when the chance of the estimated proportion differing from the actual by twice the standard error is 46 in 1,000 or .05.

6. *Measurable Characteristics* (*Variables*).

(*a*) *The Distribution of the Means of Samples.*

Sets of data or samples drawn at random from the same population will yield means of different values, varying slightly from the true mean of the whole population, owing to variations arising from sampling. These means of the samples themselves form a frequency distribution with characteristics, tending towards the normal, and a standard deviation equal to the standard deviation of the population divided by the square root of the number of observations in each sample.

Where σ_p = the standard deviation of the population
n = the number of observations in the sample.

Then the mean of the samples will be distributed with a standard deviation of $\frac{\sigma_p}{\sqrt{n}}$

Where the standard deviation of the whole population is not known the standard deviation of the distribution of the samples can be estimated from the standard deviation and number of observations in the samples under investigation. This estimate is termed the standard error of the mean.

Where σ = The standard deviation of the sample
n = The number of observations in the sample.

Then the standard error of the mean

$$= \frac{\sigma}{\sqrt{n}} \text{ and the probable error } = \frac{.67449\sigma}{\sqrt{n}}$$

As previously indicated the reliability of the sample varies with the square root of the number of items it contains.

For example, assume that the data in Example 44 is a sample. It is desired to find the mean of the population.

Example 50.

When $\left.\begin{array}{l} a = 10.24 \\ \sigma = 3.48 \\ n = 79 \end{array}\right\}$ in relation to the sample (Example 44).

Then the standard error of the mean $= \frac{\sigma}{\sqrt{n}}$

$$= \frac{3.48}{\sqrt{79}} = .391$$

When the mean of the population $= a_p = a \pm \frac{\sigma}{\sqrt{n}}$

$$= 10.24 \pm .391$$

(*b*) *The Significance of the Difference Between Means of Samples.*

Two or more samples may give different means, when it is necessary to determine whether the difference between means is statistically significant—whether the difference is attributable to:—

(1) the hazards of random sampling, or

(2) the fact that the samples may be drawn from different populations.

These questions may be determined by calculating the standard error of the difference between means.

Let σ_1 = the standard deviation of one sample;
n_1 = the number of items in the sample;
σ_2 = the standard deviation of another sample;
n_2 = the number of items in the other sample.

Then the standard error of the difference between means

$$= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

And the probable error of the difference between means

$$= .67449\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

If the difference between the two means is greater than twice the standard error of the difference between means (or 3 times the probable error of the difference between means), it is probable that they are not drawn from the same population. Conversely, if the difference between the two means is less than twice the standard error of the difference between means (S.E. diff), (or 3 times P.E. diff), then it is highly probable that the means have been drawn from the same population, i.e. that they are comparable and the variation in the means of the samples is attributable to random sampling.

For example, it is desired to ascertain whether the differences between two means derived from samples is attributable to the hazards of random sampling or to the fact that they are drawn from different populations.

Example 51.

Data	Sample (1)	Sample (2)
a	10	8
σ	2.24	2.19
n	20	20

The difference between means $= a_{\text{diff}} = 10 - 8 = 2$

The standard error of the difference between means $= \sigma_{\text{diff}}$

$$= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{5.0}{20} + \frac{4.9}{20}} = .7$$

and the probable error of the difference between means = .47.

$$\text{Then } \frac{a_{\text{diff}}}{\sigma_{\text{diff}}} = \frac{2}{.7} = 2.85 \text{ and } \frac{a_{\text{diff}}}{\text{Pe.}_{\text{diff}}} = \frac{2}{.47} = 4.26.$$

The difference between means is greater than twice the standard error (or three times the probable error) of the difference between means and therefore the samples are unlikely to have been drawn from the same population (a chance of 1 in 20) and are not comparable.

7. *Other Tests of Statistical Significance.*

Other tests of statistical significance have been evolved. These are of a more advanced nature and involve the calculation of different factors characteristic of the data under consideration and comparison of these factors with calculated, theoretrical factors appropriate to the investigation.

8. *Other Types of Curves of Frequency Distributions.*

There are other types of curves used in connection with frequency distributions. Each has its own special features and uses, but the normal curve is the most important type and, as indicated above, it forms a convenient basis for the investigation of statistical data.

Chapter XII

MUNICIPAL STATISTICS

Object: Examples: For Information of Ratepayers—of Chief Officials—The Treasurer—The Medical Officer (including Estimates of Population)—and the Surveyor: As an aid to Supervision—The Treasurer—The Surveyor: In connection with Investigations—Examples.

THE objects served in the preparation of municipal statistics can be set out under four main headings, viz.:—

To provide information—

(1) to ratepayers;
(2) to chief officials;
(3) as an aid to supervision;
(4) in connection with financial and other investigations and for the information of committees.

The general questions arising in connection with these four headings are here considered, and suggestions are made as to the information that may be supplied in each case. The actual information furnished is entirely a matter of local practice and is largely governed by the personal opinions of chief officials, members of the council, or ratepayers' associations and similar organisations.

1. *Statistics provided for the information of Ratepayers.*

The objects in supplying information under this heading may be stated as—

(*a*) To stimulate interest in local affairs.
(*b*) To supply general information to the ratepayer who is casually interested in local matters.
(*c*) To supply information to the ratepayer who displays particular interest in the business of the local authority and in local conditions generally.

The objects under (*a*) and (*b*) may be achieved by the issue of an epitome containing—

(1) Simple statements of fact, e.g.—
    The area of the district.
    The population, at the last census and as estimated to date.
    The density of the population.
    The number of inhabited houses.
    The average number of persons per house.
    The birth rate.
    The death rate.

The rateable value.
The rateable value per head of population.
The net loan debt, suitably summarised.
The net loan debt per head of population.
The net loan debt per £ of rateable value.
The average rate of interest on loans.
The assets, suitably summarised.
The assets per head of population.
The assets per £ of rateable value.
The product of a 1d. rate.
The rate poundage.
The rate poundage divided under heads of services.
The mileage of roads.
The number of people employed.
A list of the services performed, e.g. public health—tuberculosis, disinfection, maternity and child welfare, dental clinic, sanitary inspection, &c.
The charges for services, e.g. gas, water, electricity.
The cost of services, e.g.
    Public lighting—per lamp.
        —per mile of road.
    Fire brigade—per 1,000 population
        —per £100 rateable value.
    Police—per 1,000 population.
    Education—per scholar in average attendance.
The general development of services or undertakings, e.g. the development of the electricity undertaking shown by the units produced during the past (say) 15 years.
The revenue account } In a summarised form, figures to
The balance sheet } the nearest £, &c.

(2) The submission of information in pictorial or graphic form, designed to evoke interest and to readily supply general facts and figures. Many of the items enumerated admit of pictorial presentation:—

(*a*) The items regarding loan debt and assets may be presented in the form of block or bar diagrams.

(*b*) The rate poundage, divided under heads of services, may be shown as a block suitably sub-divided or as segments of a circle.

(*c*) The rateable value over a number of years may be illustrated by a graph or a bar diagram.

(*d*) The general development of, for example, the electricity undertaking may be shown in the form of a graph, the curve of which is supported at annual points by pylons. The latter, although unnecessary, serve to draw attention to the diagram and to the annual figures, while the trend of the 'cable' indicates the development.

(*e*) Local landmarks may be utilised to evoke interest. For example, an illustration may be given of the town clock together with a statement to the effect that for every hourly chime the council spends so much on health services, education, &c.

In the case of (*c*) the more detailed statements of the abstract of accounts, the reports of the medical officer of health and other chief officers, the general report of the council, and similar publications are required.

The notes following give some indication of the details usually shown in the abstract of accounts and the report of the medical officer of health. They are not intended to be exhaustive, but merely to draw attention to the type of factors necessary for an appreciation of local affairs and the use that may be made of diagrams designed to assist in the assimilation of the import of the facts.

(1) The abstract of accounts should be designed to show, in addition to the financial items suggested for an epitome, greater detail generally, an especially in regard to the following subjects, viz.:—

(*a*) The income and expenditure of the different services, trading undertakings, and funds, shorn of unnecessary detail, but giving sufficient information to be of value to an interested inquirer.

(*b*) Similar considerations apply to the balance sheets of the various funds and undertakings, and to the aggregate balance sheet.

(*c*) Detailed statements of the—

Loans outstanding.
Capital outlay discharged.
Provision for the redemption of debt.
Assets.

These items may be classed under the headings of—

Productive.
Non-productive.
Realisable.
Non-realisable.

(*d*) A statement could also be prepared of the loans outstanding and the years in which they fall due for redemption.

(*e*) Detailed statements of the working of trading undertakings, e.g. electricity—comparative figures showing the units generated, units sold, coal used, &c.

The trend of events in regard to these items readily admits of pictorial illustration, especially by the use of bar diagrams and graphs.

For example, the trend of the policy in regard to the financing of capital expenditure may be depicted by plotting the items, over a number of years, indicated in paragraph (*c*), in the form of a graph or a series of graphs, or bar diagrams.

(2) In the case of the report of the medical officer of health, comparative details will be shown of the vital statistics of the area, the progress of the various services, the sanitary supervision, and the housing conditions. The following brief notes give some indication of the material usually supplied:—

(1) Population.
  (*a*) The population as at the last census.
  (*b*) The estimated population to date.
  (*c*) The density of the population.

(2) Births, deaths and marriages.
  (*a*) The actual figures.
  (*b*) The rate per 1,000 of population.
  (*c*) In the case of the first two items—
    (i) The numbers subdivided according to sex.
    (ii) The figures adjusted for inward and outward transfers.
  (*d*) The infant mortality rate.
  (*e*) The maternal mortality rate.
  (*f*) The number of deaths under the headings of the various causes.
  (*g*) The number of deaths under the headings of sex, age and cause.

(3) Diseases.

Special reports and figures in regard to the incidence of certain diseases and the services undertaken regarding them, e.g.—

Tuberculosis.
Cancer.
Smallpox.
Measles, and other infectious diseases.

(4) Miscellaneous services.

Special reports and figures in regard to the miscellaneous health services of the area, e.g.—

Vaccination.
Disinfection.
Bacteriological examination.
Maternity and child welfare.
Infant life protection.
Food and drugs.
Mortuary.

Smoke abatement.
Destruction of rats and mice.

(5) Sanitary conditions.

Special reports and figures in regard to the sanitary conditions of the area, e.g.—

Inspections made.
Defects remedied.
Prosecutions.
The general character of the district from a sanitary point of view.

(6) Housing.

Special reports and figures in regard to housing, overcrowding and rehousing, e.g.—

The number of inhabited houses and the average number of persons per house.
Particulars of overcrowding.
Particulars of the abatement of overcrowding.
Particulars of rehousing undertaken by—
  (*a*) the local authority;
  (*b*) private persons.

As in the case of the abstract of accounts, these facts admit of ample opportunity of illustrating the trend of the various factors by the use of graphs and block or bar diagrams, and so adding to the utility of the data.

2. *Statistics Provided for the Information of Chief Officials.*

The preparation of statistical matter is of importance to all chief officials as an aid in the administration of their departments.

Statistical data prepared under this heading is designed to present the salient facts of local matters and their trend.

Examples of the type and nature of the special information required, apart from the general information contained in the epitome, is given in regard to the departments of the treasurer, the medical officer and the surveyor.

The graphic method of illustrating facts and their general trend may be employed, with advantage, to many of the items indicated in the following paragraphs.

(1) *The Treasurer.*

The data would largely be an extension of that supplied in the abstract of accounts, and would include some or all of the following:—

(*a*) *Information to Indicate the Trend of Development.*

Separate tables to give detail and a composite table showing comparative figures over a number of years in regard to the following matters:—

- (i) The area.
- (ii) The population (the census figures and as estimated to date).
- (iii) The population per acre.
- (iv) The rateable value.
- (v) The rateable value per head of population.
- (vi) The product of a 1d. rate.
- (vii) An analysis of the rateable value into classes, e.g. dwelling-houses, shops, garages, industrial properties, &c., and into groups, e.g. not exceeding £10, exceeding £10 and not exceeding £20, &c.

(*b*) *The Working Balance.*

A comparative table to show—

- (i) The rate income required and rate poundage.
- (ii) The rates raised and rate poundages levied.
- (iii) The surplus or deficiency on the year's working, i.e. the amount added to or taken from the balance.
- (iv) The cash balance brought forward and carried forward.

(*c*) *The Revenue Accounts.*

- (1) An analysis of the various revenue accounts including trading undertakings, suitably subdivided under the headings of services, votes, &c., over a number of years.
- (2) Amplifications of abnormalities arising from item 1.

(*d*) *The Assets.*

An analysis of the capital accounts, over a series of years, showing services, trading undertakings, &c., and summarised under the headings of realisable and non-realisable, remunerative and non-remunerative assets.

(*e*) *The Net Loan Debt.*

An analysis suitably subdivided under the headings indicated in item (*d*).

(*f*) *The Loan Debt.*

- (1) An analysis of the loans outstanding, classified under the headings of—
  - Stock.
  - Mortgages.
  - Short term loans.
  - Overdrafts, &c.,

  and further subdivided under the rates of interest payable.

(2) A table showing the amounts becoming due for repayment, analysed under the headings of the year in which repayable, and also the rate of interest.

(*g*) *Special Funds.*

An analysis, over a period, of the transactions of special funds, e.g.—

(1) Insurance funds.
(2) Superannuation funds.
(3) Charities.
(4) Reserve funds.
(5) Loans funds.
(6) Sinking funds.

(2) *The Medical Officer.*

(*a*) The information required would be an extension and amplification of that supplied in the annual report.

(*b*) Notes in regard to some of the 'rates' usually supplied in the annual report are given below:—

✓(1) *The Birth Rate.*—This is calculated by—

(*a*) Ascertaining the number of live births recorded, in each year, in the district, subject to adjustment for inward and outward transfers.
(*b*) Multiplying the figure so obtained by 1,000.
(*c*) Dividing the result of (*b*) by the estimated mid-year population of all ages of the district.

The result is the birth rate per 1,000 of population and is usually given to one place of decimals.

✓(2) *The Death Rate.*—This rate is calculated in a similar manner to the birth rate and is the total number of deaths recorded in the district, subject to adjustment for inward and outward transfers to place of residence, per 1,000 of the estimated mid-year population of the district:—

Specific death rates refer to—

1. Deaths analysed under causes.

2. Deaths analysed according to age, sex, occupation, &c., and are given in terms per 1,000 of the estimated mid-year population of the district in respect of item 1, and per 1,000 of the estimated mid-year population of the group in respect of item 2.

✓(3) *The Marriage Rate.*—This may be ascertained by multiplying the number of marriages by two and proceeding as indicated above.

✓(4) *The Infant Mortality Rate.*—This rate is calculated in a similar manner to the birth rate, subject to the undermentioned differences:—

(*a*) An infant refers to a child under one year of age.

(*b*) The rate is based on the number of deaths among infants expressed in terms of 1,000 of the live births during the year.

(5) *Still Births* are expressed:
per 1,000 of total births,
per 1,000 of live births.

(6) *Reproduction Rates.* In this case, where factors indicative of the replacement of population are required, female children only are taken into consideration.

(*a*) *Gross Reproduction Rate.* The sum of the number of female live births per 1,000 women, for the year in question, for the 7 age groups 15 years by 5 to 45, is multiplied by 5 (the class interval of the age group) and the resulting product divided by 1,000.

(*b*) *Net Reproduction Rate.* Allowance is made for the mortality of females from birth to the age of the mother at the birth of the child. The number of female live births per 1,000 women for each age group is related to the mortality table for that group, the sum of the resulting factors is multiplied by 5 and the resulting product divided by 1,000.

(7) *Corrected Rates.*

The particulars given above refer to the crude rates. In order to make possible greater accuracy in comparison with figures for other areas, methods of 'correction' or 'standardisation' have been devised. In the case of the death rate, for example, the average mortality at different ages has a great influence on the crude rates. Local figures must, therefore, be adjusted to allow for the mortality of the population assessed under sex and age groups. The method is essentially the application of weights and in principle has been achieved as follows:—

(1) The death rate is computed for each age group under the heading of sex, e.g. (0-5 years, 5-10, 10-15, &c.) in regard to the country as a whole. This rate is termed the 'standard death rate'.

(2) The figures thus obtained are then utilised in calculating a rate for individual areas based on the sex-age-groups of the local population.
The resulting rate is termed the 'standard death rate' of the area.

(3) A correction factor is then calculated for individual areas by dividing the standard for the country as a whole by the standard for the district.

(4) The corrected death rate may then be obtained by multiplying the recorded crude rate by the correction factor.

The formula may be stated thus—

$$\text{The corrected death rate} = \text{the recorded (crude) death rate} \times \frac{\text{the general standard}}{\text{the standard for the local area.}}$$

Different methods or bases have been adopted by the Registrar-General in attempting standardisation. These are summarised below:—

Period 1911-1920.

A standardising factor was calculated by relating the average death rates, for the period 1901-10 of the country, to the local population in 1911 and to the national population in 1901 ; when the local standard death rate = the local crude death rate × the standardising factor.

Period 1921-1933.

The standardising factor was amended by utilising the average death rates for the country for 1920-22, which were related to the local population in 1921 and the national population in 1901.

Period 1934 *onwards.*

In 1934 a different method was adopted, which was designed to provide information as to the extent to which the local crude death rate is greater or less than the national rate, after taking into account the mortality of the country as a whole, analysed under sex and age groups. This method involves the calculation for each local area of

1. An areal comparability factor (A.C.F.) and
2. A standardised ratio, i.e. the ratio of the adjusted local death rate to the national crude death rate.

(These factors are given, for each district, in the Registrar-General's Annual Statistical Review).

The areal comparability factor is calculated as follows:—

1. Standard death rates of the country, at sex and age groups, were obtained by dividing the deaths in the years 1930-32 by three times the census population at 1931, of those groups.
2. Standard death rates as obtained in Item 1 were multiplied by the 1931 population (grouped) of the local area, when the sum of the products so obtained were divided by the total population of the area, giving the expected mean local death rate or local index rate.

3. The ratio of the mean crude death rate of the country (i.e. the mean of the rates in the various sex-age groups in Item 1) to the local rate (termed the 'index' rate—Item 2) gives the areal comparability factor for the area.

The adjusted local death rate = the crude local death rate (for the year in question) × the A.C.F.

The standardised ratio, mentioned above, is the ratio of the adjusted local death rate to the national death rate. This ratio provides a measure of the relative incidence of mortality between different areas. Where it is desired to make comparisons with years in which a standardised death rate (based on 1901 population) has been calculated, the adjusted local death rate is multiplied by a time comparability factor (T.C.F.). The time comparability factor is the ratio, for a particular year, of the standardised national rate to the crude national rate for that year.

(This factor also is given, for each year, in the Registrar General's Annual Statistical Review.)

The standardised local death rate = the crude local death rate × the A.C.F. × T.C.F.

The term combined comparability factor (C.C.F.) is applied to the product of the A.C.F. and the T.C.F. of the area.

Attention is also drawn to a further method of expressing mortality by a single figure, viz., the equivalent average death rate. This method is not affected by the proportions of the population in the different age groups. The equivalent average death rate is the arithmetic average of the rates at age groups of 5 years. This rate is thus equivalent to calculating a standardised rate covering all ages, but based on the population distributed equally over the different age groups. It is thus necessary to ensure that undue weight is not given to the average by the factors arising from groups of normally high mortality. The equivalent average death rate for groups of ages up to 65 has proved a useful basis for general purposes of comparison.

(*c*) The estimation of population in local areas at particular dates is a matter of difficulty and examples in regard to this problem are given below. It is generally recognised that the population varies in geometric progression, i.e. the perhaps more familiar compound interest calculation, and the following examples are based on this assumption.

Method (1). This method assumes that the rate of increase, i.e. births plus immigrants less deaths and emigrants will be the same as in the previous intercensal period.

Let p = the population at the last census.
r = the annual rate of increase.
n = the period elapsing from the date of the last census to the date on which the estimate is required.

Then the estimated population $= p\ (1 + r)^n$

Example 52.

$p = 101{,}147$
$r = .0032 = 3.2$ per thousand
$n = 4$ years.

Applying the formula.

$$\text{The estimated population} = \left\{\frac{1{,}003.2}{1{,}000}\right\}^4 \times 101{,}147$$
$$= 102{,}450.$$

Method (2). This method assumes that the average population per inhabited house is the same during both periods. Ascertain—

(*a*) the average population per inhabited house as at the last census;

(*b*) the number of inhabited houses at the date upon which the estimate of population is required. (This information may be obtained from the rate books.)

Example 53.

Where (*a*) = 3.013
(*b*) = 35,171
The estimated population = 35,171 × 3.013
= 105,970.

Method (3). Ascertain—

(*a*) the population at the last census;
(*b*) the death rate (from last census to date of estimate);
(*c*) the birth rate (from last census to date of estimate);
(*d*) the number of inhabited houses at the last census;
(*e*) the number of inhabited houses at the date of the estimate.

Example 54.

Where (*a*) = 101,147
(*b*) = 12.3 per 1,000 } difference = 3.2 per 1,000
(*c*) = 15.5 per 1,000 } (increase).
(*d*) = 34,011 } difference = 1,160 (increase).
(*e*) = 35,171 }

The factor of the average number of persons per house arising in connection with the increase of inhabited houses is entirely one for the application of local knowledge, and the accuracy of the estimated population will depend largely upon the accuracy of this factor and its relative importance.

Assume this average to be 2.6.

Then the estimated population (4 years from the date of last census, $= \left\{\frac{1{,}003.2}{1{,}000}\right\}^4 \times 101{,}147 + (1{,}160 \times 2.6)$

$= 105{,}460.$

Note.—In the foregoing examples it has been assumed that the population increases. The principles, however, arising in the calculations are similar where the population is decreasing.

(3) *The Surveyor.*

Here again the general information would be as contained in the epitome, amplified in regard to the items of special interest to the surveyor. Examples of the detailed information usually called for are here given.

(*a*) *Information to indicate the trend of Development.*
Similar information to that indicated under the heading of the treasurer, but with further information in regard to the bearing of town planning on the subject.

(*b*) *Maintenance.*
(1) *Highways.*—A comparative statement extending over a period of years and analysed under road classifications, and, if required, further subdivided under individual roads, giving particulars of—
  (*a*) the type of foundation;
  (*b*) the type of surface;
  (*c*) the nature and date of repairs and resurfacing carried out;
  (*d*) the cost of these works.
(2) *Housing.*
Tables giving details in regard to the following:—
  (*a*) date of erection;
  (*b*) types erected;
  (*c*) initial cost, analysed under buildings, land, roads, sewers;
  (*d*) general particulars of construction;
  (*e*) repairs—
    (1) annual cost;
    (2) nature—drawing attention to special items.
(3) *Miscellaneous Properties.*—Similar tables to those indicated under the heading of housing with additional information as to special items, e.g. extensions.

(4) *Plant.*—Details of fixed and movable plant, vehicles, &c., showing—
  (*a*) Date of purchase;
  (*b*) Repairs and renewals—annual cost and nature.

3. *Statistics Provided as an Aid to Supervision.*

There is no clear dividing line between information supplied under this heading and that under heading No. 2, a broad generalisation being that this heading covers the suitable summarisation and presentation of current detail, while the former has a wider scope. With a few exceptions the items will thus be the periodical revision of the annual figures already indicated. Examples of information designed to assist in supervision, which have not been previously indicated, are given in respect of the treasurer and the surveyor.

1. *The Treasurer.*
  (*a*) *Statement of Rates Collected.*
    (1) The weekly amounts collected may be plotted as a graph or the cumulative amounts as an ogive. Should the latter method be adopted, a line could also be plotted showing the rate requirements.
    (2) For comparisons with previous years the relevant figures may be reduced to percentages and the percentages plotted as before.
  (*b*) *Income of Trading Undertakings and Miscellaneous Income.* The weekly income over a period may be plotted in the form of graphs, with separate curves for each year.
  (*c*) *The Financing of Capital Expenditure.*
    (1) Curves plotted of the market prices of the chief government securities and the stocks of the local authority will show at a glance the trend of the market and provide valuable information as to the issue and purchase of securities, rates of interest and other aspects of borrowing.
    (2) Statements or diagrams showing the amounts and proportions of funded and unfunded debt.
  (*d*) *Routine Work.*—Much can be done in regard to internal organisation where the cost of routine work is known and a suitable set of 'unit costs' compiled.

  The basis in such an inquiry is the mode, i.e. the analysis of the 'cost per job' into factors of the normal or the most predominant conditions, but the normal should be suitably weighted to allow for the difficulties that arise from time to time in the actual practice.

  Examples indicating the nature of these inquiries are given below.

Item	*The cost of and time required in performing the following operations.*
1. Dividend Warrants, Rate Demand Notes, Gas and Electricity accounts, &c.	(*a*) Addressographing, calculating, typing, despatching. (*b*) Posting debits.
2. Income.	(*a*) Issuing and despatching receipts. (*b*) Posting credits.
3. Wages.	(*a*) Checking, preparing and making up pay packets. (*b*) Paying wages.
4. Housing Rents, Slot Meters, &c.	(*a*) Actual collection. (*b*) Posting credits.
5. Rates.	(*a*) Actual collection. (*b*) Posting credits. (*c*) Making adjustments for broken periods, &c.

2. *The Surveyor.*
   - (*a*) *The progress of work.*
     - (1) Statements in the nature of cost sheets showing the periodical progress of work undertaken.
     - (2) Statements comparing the actual progress of work with the time table or programme.
   - (*b*) *Unit Costs.*—As indicated under the heading of the treasurer, the comparison of actual with scientifically prepared unit costs forms a useful aid to control. Examples of units applicable are given below.
     - (1) Sewers—per mile.
       per 1,000 of population.
     - (2) Cleansing—
       - (*a*) Refuse—Collection } per ton.
         Disposal } per 1,000 of population.
       - (*b*) Street cleansing—per mile.
         per 1,000 of population.
       - (*c*) Gully emptying—per gully.
         per 1,000 of population.
         per mile of road.
     - (3) Highways—Maintenance, subdivided under the headings of type of surface and foundation and width of road—
       per mile.
       per 1,000 of population.

(4) Lighting—per lamp.
per mile of road.
per 1,000 of population.

(5) Vehicles—analysed under type and capacity—
per mile.
per ton.
per ton mile.

4. *Statistics Provided in Connection with Financial and other Investigations, and for the Information of Committees.*

The question of the collection and preparation of data assumes great importance when considered in connection with its aspect relating to investigations. Notes on the information required in connection with certain investigations are given below.

(1) *Compounding in connection with Rateable Properties.*—In this case it is desired to investigate:—

(*a*) The need for compounding.
(*b*) The financial effect of compounding.

Tables could be prepared showing an analysis of rateable properties under the following headings:—

(1) The number within and outside the compounding limits.
(2) The value within and outside the compounding limits.
(3) The percentages relative to item (2).
(4) The arrears within and outside the compounding limits.
(5) The percentages relative to item (4).
(6) The cost of compounding in terms of 1d. rate, i.e. the reduction of the product consequent upon the introduction of compounding.
(7) The estimated increase or reduction in the cost of collection arising from such an alteration.

From the foregoing the necessity for, and financial effect of, the suggested change can be ascertained.

(2) *Compounding for Stamp Duty on Stock Transfers.*—It is an open question whether it is more economical to compound for stamp duty on stock transfers or to pay the duty as it arises. The problem is largely one of local application, but much use may be made of past experience in deciding future policy.

To assist in a decision, information under the following headings could be prepared:—

(1) The period of the stock.
(2) The amount of the issues.
(3) The charge for composition of stamp duty.
(4) The amounts of stamp duty payable on actual transfers.

(3) *The Purchase of a privately owned Trading Undertaking.*—It is desired to ascertain:—

(1) A normal year's profits.
(2) The capacity of the undertaking.
(3) The condition of the plant, both in regard to repair and obsolescence.
(4) The future trend of the undertaking.

The purchase price would in all probability be based on the estimated normal profits of the undertaking, capitalised on an agreed period of years, subject to adjustments arising under items (2), (3) and (4). Indications of the nature of the information required are given below. To save too much generalisation it is assumed that the undertaking supplies gas or electricity.

(1) The investigation under item (1) would comprise the information indicated below, viz.:—

  (1) The Capital Account.—A detailed analysis of the capital account over a number of years showing the nature and purpose of the expenditure.

  (2) The Revenue Account.—A detailed analysis of the revenue account over a number of years showing:—
    (*a*) the working expenses;
    (*b*) the income;
    (*c*) the profit and its appropriation;
    (*d*) the units produced and sold;
    (*e*) unit costs of the above items, where applicable;
    (*f*) detailed information of any exceptional or abnormal items arising under the above headings.

(2) and (3) The investigations under items (2) and (3) are largely questions of fact supported by expert opinion and would doubtless be placed in the hands of technical advisers.

(4) Information on the lines indicated below would be of assistance in investigating future prospects, viz.:—

  (1) Tables extending over a number of years with, say, monthly totals of—
    (*a*) the units produced;
    (*b*) the units sold.

  These tables would indicate the trend of the undertaking and draw attention to seasonal or other fluctuations.

  (2) Tables analysing the consumption over a period under the following headings:—
    (*a*) units consumed, i.e. class interval;
    (*b*) number of consumers;
    (*c*) units consumed per class, i.e. (*a*) × (*b*).

  These tables will decide the 'model' consumer and the general sources of the income of the undertaking.

(3) Tables analysing the consumption over a period under the heading of tariffs, viz.:—
  (*a*) units consumed, i.e. class interval;
  (*b*) number of consumers;
  (*c*) units consumed in regard to each method of charge, i.e. (*a*) × (*b*).

  Other items requiring consideration would be the trend of the population and the possibility of competition.

The above information (Item 4) is by no means exhaustive and in any case it would vary with the type of undertaking under consideration.

From it, however, it would be possible to investigate the possibilities of increasing the consumption by endeavouring, for example, to—
(1) eliminate seasonal fluctuations, or by
(2) extending the model class.

Further, the figures would afford some assistance as to how this end could be achieved, e.g. by—
(1) a revision of prices;
(2) a revision of tariffs;
(3) an extension of the methods of inducing consumption, e.g. 'free' (rental) wiring or a reduction in the price of hiring appliances.

Chapter XIII

MECHANICAL AIDS TO THE PREPARATION OF STATISTICAL DATA

Advantages—Machinery—Adding—Listing—Calculating—Sorting and Tabulating: The Card System.

THE advantages accruing when mechanical aids are sought to the preparation and presentation of statistical data may be summarised as follows:—

1. *Speed.*—The value of the use of statistics is greatly increased where the data is presented within a short period of its arising.
2. *Accuracy.*—The partial elimination of the 'human element' arising in complex calculations or tedious operations is a great aid to accuracy.
3. *Detail.*—Far greater detail and a greater number of combinations of data are available when use is made of mechanical appliances.

Notes on the nature and capabilities of the various appliances available are given below:—

(1) *Adding Machines.*
  (*a*) To give a total.
  (*b*) To give sub-totals and a cumulative total.
(2) *Listing Machines.*
  (*a*) To list items and give a total.
  (*b*) To list items and give sub-totals and a cumulative total.
  (*c*) To admit of the above operations under various headings, i.e. providing several columns of figures.
(3) *Calculating Machines.*
  (*a*) To add.
  (*b*) To subtract.
  (*c*) To multiply.
  (*d*) To divide.
  (*e*) To raise to a power.
  (*f*) To extract a root.
(4) *Sorting and Tabulating Machines.*
  The operations arising under this heading consist of:—
  (1) *Recording.*
    (*a*) The numerical value of the information is recorded by a punching machine which cuts holes in specially prepared cards. The values of the holes are previously arranged according to a defined coding system.
    (*b*) The cards are then passed through a device for purposes

of checking. The operation is similar to the punching, with the exception that no holes are cut and that the machine stops where the card differs from the data indicated.

(2) *Classifying.*
Classification is also achieved by punching holes in the cards, under the headings of code numbers.

(3) *Sorting.*
The cards are passed through a sorting device which selects the cards under the code number required from the other cards in the batch.

(4) *Tabulating.*
The selected cards are passed through the tabulator which converts the punched holes in the cards into numbers and prepares a total or a list and total(s).

Machines capable of performing the operations of adding, listing and calculating, are comparatively reasonable in price and are installed in the offices of many local authorities.

Sorting and tabulating machines, however, are more expensive and are usually found only in the offices of the largest authorities.

Where it is found that a punched card system is desirable, but that expense debars the procuring of these machines, it may be feasible to let the work out under contract.

Where desired, however, the card system may be applied in a modified form. The principle is the same as in the case of the most efficient and expensive machinery indicated under heading (4) and may be achieved as indicated below:—

(1) *Recording.*—The cards are supplied with holes punched near the edges. The numerical value is recorded by cutting away (with a multiple key punch or an instrument similar to a railway ticket cutter), the portions of the cards between the holes and the outer edge in respect of the values (i.e. the perforated holes) required.

(2) *Classifying.*—The classification is similarly recorded.

(3) *Sorting.*—The cards in respect of separate code numbers are extracted by passing a rod through the hole applicable. By this means the cards required drop away from the rod. Any other cards in the batch are retained on the rod owing to the fact that the portion of the card between the hole and the outer edge is not cut away.

(4) *Tabulating.*—The values indicated on the cards required are then converted into numerical values by the operator and listed and tabulated on a listing machine.

Note.—(1) Coding may be simplified or extended by the use of cards of differing colours.

(2) To facilitate ensuring that the cards are the correct way up a corner, say the top right hand, may be cut away.

Chapter XIV

PUBLISHED STATISTICS

Guide to Current Official Statistics—Official Statistics—Other Statistics.

IT IS only possible within the confines of this book to give very brief references in regard to some of the published statistics available, and of interest to local government officers.

The amount of statistical information prepared by government departments, for example, is enormous, as is indicated by the fact that the 'Guide to Current Official Statistics' consists of some 350 pages of printed matter.

The Guide referred to is prepared by the Permanent Consultative Committee on Official Statistics. It gives comprehensive references to the published information prepared by the various government departments, suitably analysed and cross-referenced.

The contents of some of the chief statistical publications from a local government standpoint, are indicated in the following pages.

Central Statistical Office.

Monthly Digest of Statistics.

Prepared by the Central Statistical Office in collaboration with the Statistics Divisions of Government Departments. The figures given are in respect of the years 1935 to 1944, and monthly since January, 1944. The first issue of this Digest was made in January, 1946.

'The aim is to publish in summary form statistics which show the changing pattern of the economic activity of the country. All statistics are, therefore, presented in the form of time series. . . . The scope of the Digest will be revised from time to time to include new statistical information as it becomes available.'

The January, 1946, Digest contained information under the following headings:—

1. Employment.
2. Fuel and Power.
3. Raw materials.
4. Manufactured goods.
5. Building.
6. Agriculture and food.
7. External trade.
8. Merchant shipping.

9. Inland transport.
10. Finance.
11. Wages and prices.

Statistical Abstract of the United Kingdom. (Annual.)

This abstract is issued in conjunction with the Ministry of Labour and the Registrar-General. It is the most comprehensive summary issued of the statistics of the country and contains details of all statistics relating to trade and individual commodities and matters analogous thereto.

Ministry of Agriculture and Fisheries.

(*a*) The Journal of the Ministry of Agriculture. (Monthly.)

Number of students attending agricultural classes—over a number of years.

Grants to institutions and local authorities over a number of years.

Courses provided by local authorities.

Miscellaneous statistics regarding agricultural education and research.

Crops—suitably subdivided,
    average prices and index numbers,
    average acreages under cultivation,
    estimated produce,
    estimated yield per acre,
    subsidies granted.

National Marks—miscellaneous statistics.
Milk grading ,, ,,
Foot and mouth disease—particulars of outbreaks.
Employment—casual, regular, male and female,
    —wages, rates.
Live Stock—miscellaneous figures suitably analysed,
    —imports and exports,
    —miscellaneous figures re improvement of stocks.

(*b*) The Agricultural Market Report. (Weekly.) Miscellaneous statistics dealing with the market aspect of agriculture.

(*c*) Agricultural Statistics. (Annual.)

Part 1.—Detailed statistics analysed under the undermentioned general headings:—
    Acreages under cultivation,
    Production of crops,
    Live stock,
    Miscellaneous summaries.

Part 2.—Prices of Agricultural Produce,
    Supplies ,, ,,
    Miscellaneous statistics regarding agricultural produce.

Colonial Office.

Colonial Reports. (Annual.)

Separate publications for the various Colonies, giving detailed reviews and statistics in regard to the following:—

- Local Administration
- Finances
- Trade
- Agriculture
- Health
- Vital statistics
- Justice
- Education
- Meteorological reports.

Board of Control (Lunacy and Mental Deficiency).

(*a*) Annual Report.

Part 1. Classified list of patients—private, rate aided, criminal, temporary, certified, voluntary; sex, type of institution and by whom provided. Statistics as to patients—admissions, discharges, transfers, deaths, diseases. Accommodation available.

Part 2. General summary of Research work.
Miscellaneous statistics as to accommodation available, &c.

(*b*) Costing Returns. (Annual).
(*a*) County and borough mental hospitals.
(*b*) Certified institutions for mental defectives.
Miscellaneous summaries as to the upkeep, &c., of institutions and inmates.

Ministry of Education. (Annual.)

(*a*) Annual Report.

Adult Education
Evening Institutions } types of courses, local authorities, subjects, attendances.

Blind, Deaf and Defectives—number, accommodation, types of schools, ages of scholars, &c., subjects, &c.

Nursery Schools
Play Centres } number, average attendances, local authorities.

Schools—number, type, &c., suitably analysed.

Scholarships—awarded, held, entrants, taken up, declined, type, sex, maintenance allowances.

Pupils—in average attendance, detailed analysis.

Teachers—numbers, subjects, full time, part time, pensionable, training, &c.

Authorities—types, areas, populations.

School Medical Service inspections made, defects, treatments, &c., with detailed summaries.

Nautical Training—ships and schools, number, pupils, &c.
Pensions—number, values, types, sex, contributions.

(*b*) Memorandum on Board of Education Estimates.

Expenditure by the Board Expenditure by Local Authorities	Elementary and Higher, analysed under main headings, estimated and actual—per pupil (detailed).

Capital Expenditure—approved, expended, loans sanctioned —raised, outstanding, charges.

Electricity Commission. (Annual.)

(*a*) Annual Report.

Generating Stations—number, authority, company, other undertakings—units generated
Units generated (totals)
Fuel used
Consumption of electricity
Particulars as to transmission lines
Accounts of the income and expenditure of the Commissioners
Miscellaneous statistics as to—
- loans sanctioned
- output
- generation
- consumption
- expenditure

summarised in respect of the various undertakings.

(*b*) Generation of Electricity.

Detailed analyses and summaries in respect of
- units generated
- fuel consumption
- thermal efficiency
- maximum load
- load factor.

(*c*) Engineering and Financial Statistics.

Detailed analyses and summaries in respect of—
- loans sanctioned
- loans raised
- net debt
- capital expenditure
- income from sale of current
- total units sold
- working expenses
- appropriation of surplus
- system of supply

plant
maximum load
load factor
consumption of fuel
units generated, purchased, used at works, sold, unaccounted for
number of consumers
populations of supply areas.

General Register Office.

(*a*) Weekly Returns of Births and Deaths, &c., registered in county boroughs and other large towns, and certain infectious diseases notified in boroughs, urban and rural districts.

(*b*) Quarterly Return of Births, Deaths and Marriages and certain infectious diseases—county boroughs, urban and rural aggregates of administrative counties.

(*c*) Census of England and Wales, 1931.

Preliminary Report.

Population of each administrative county, county borough, metropolitan borough, municipal borough, urban and rural district—electorate, sex—comparative figures over a number of years.

(*d*) County Volumes.

Population of administrative county and local areas within the county—in each ward or parish, sex, ages, marital condition, electorate—comparative figures.
Structurally separate dwellings
Number of occupying families
Number of rooms occupied, number of rooms occupied under number of occupying families per house, number of rooms in dwellings.
Sizes of families, rooms occupied, average numbers per room.

(*e*) Registrar Generals Statistical Review of England and Wales.

Part 1. Medical.

Climatic conditions
Births and birth rates (analysed)
Maternal mortality rates (analysed)
Infant mortality rates (analysed)
Deaths and death rates—sex—age—crude—standardised—causes—areas.
Statistics of various diseases
Marriages and marriage rates
Population, &c.

Part 2. Civil.

- Births and birth rate
- Marriages and marriage rate
- Deaths and death rate
- Divorces
- Electorate
- Infant mortality
- Migration
- Population

Ministry of Health.

(*a*) Annual Report.

A summary of the activities of, and supervised by, the Ministry, including—

- National health insurance
- Welfare of the blind
- Population
- Death and mortality rates
- Birth rates
- Public assistance
- Food and drugs
- Progress of housing and building costs—state assisted schemes—small dwellings, &c., advances—slum clearance—exchequer contributions.
- Infectious diseases
- Summaries of income and expenditure of local authorities
- District audit
- General exchequer grant
- Loans sanctions and loan debt outstanding
- Distressed areas
- Local taxation licences
- Maternity and child welfare
- Mental deficiency
- Superannuation of local government officers
- Widows, Orphans and Old Age Contributory Pensions
- Old age pensions
- Rating, rateable values, rates levied
- Miscellaneous health services
- Refuse collection
- Sewage disposal, &c.
- Town and country planning
- Water supply
- Miscellaneous services—tuberculosis treatment and care, milk grading, vaccination, &c.

(*b*) Financial Reports. (Annual.)

Statement showing particulars as to Rates and Rateable Values in respect of Boroughs, Urban and Rural Districts in England and Wales—

(1) The rate poundage.

(2) The rateable values.

(3) The rates collected.

(4) The rates collected per head of population.

Local Government Financial Statistics—England and Wales (Local Taxation Returns). General information as to each area. Figures, suitably analysed, showing all sources of income and expenditure and particulars of loans outstanding, analysed over services, and summarised, and certain other statistics, e.g. rateable values, Government Grants, particulars of Differential Rates (where existing), &c.

Part 1 In respect of Public Assistance (Poor Law) Authorities.

Part 2 In respect of the London County Council, the City of London, the Metropolitan Borough Councils, the Metropolitan Police, the Metropolitan Water Board, and the County Borough Councils.

Part 3 In respect of the County Councils, the Municipal Borough Councils, the Urban and Rural District Councils, and Joint Boards, &c., other than the authorities shown in Part 2.

Railway Assessment Authority and Anglo-Scottish Railways Assessment Authority.

Separate reports of the proceedings of the two authorities together with financial transactions.

Accounts of the National Health Insurance Funds.

The receipts and payments of the funds, together with the report of the auditor general.

(*c*) Costing Returns. (Annual.)

Giving statistics as to the average number of beds provided and occupied, the average staff, and the average weekly cost per inmate, suitably analysed. In respect of:—

Part 1 Poor Law Hospitals, General Hospitals under the Control of Local Authorities, Sanatoria and Maternity Homes.

Part 2 Poor Law Institutions.

Part 3 Poor Law—Children's Homes.

(*d*) Public Assistance.

Poor Relief, England and Wales. (Quarterly.)

Number of persons in receipt of relief—institutional—

domiciliary—medical—rates per 1,000 of population—casuals.

See also 'Costing Returns'.

(*e*) The State of Public Health, Annual Report of the Chief Medical Officer of the Ministry of Health.

Comprehensive details of the health services, vital statistics, &c., of the country.

Home Office. (Annual.)

(*a*) Road Accidents caused by Vehicles and Horses.

Miscellaneous statistics as to number of accidents—
deaths
injuries

(*b*) Police, Counties and Boroughs.

Population of police areas.
Pensions—expenditure—gratuities—pensions—allowances.
Contributions towards pensions.
Strength
each rank
daily average.
special constables
women police.

(*c*) Report of the Commissioner of Police for the Metropolis.

Similar statistics to those indicated above with details as to crime, lost property, fingerprint registration, probation of offenders, &c.

(*d*) Metropolitan Police Fund.

Accounts of the income and expenditure, suitably analysed, together with detailed summaries as to pensions, &c.

Board of Inland Revenue.

Annual Report of the Commissioners of H.M. Inland Revenue.

Total number of Registered Companies.
Total number of Limited Partnerships.
Particulars, suitably analysed, of Estates liable to Estate Duty.
Assessments to Income Tax—gross income, net income (each schedule), allowances, reductions, &c.
Particulars of sources of incomes, e.g. British, Dominion and Foreign Governments, &c.
Assessments to Sur Tax—number of assessments, and amounts.
Receipts from the various Duties—income and sur-tax, stamp duties, estate duties, &c., detailed analysis.
Rates of Income Tax.
Land Tax—general particulars.

Ministry of Labour.

Ministry of Labour Gazette. (Monthly.)

- Particulars as to accidents (industrial), fatal, non-fatal.
- Bank rate, changes over a number of years.
- Three months bills—average rates.
- Building progress—estimated cost.
- Coal—estimated exports.
  - —estimated output.
  - —cost of production—summary.
- Cotton—average manufacture.
- Diseases—industrial summaries.
- Employment—summaries—coal, iron and steel, tin, cotton, engineering, shipbuilding, printing, docks, shipping, fisheries, railways—hours of labour, general statistics of employment exchanges.
- Exports—all articles—summaries.
- Imports ,, ,,
- Factories and workshops—number registered, analysed.
- Index number of industrial production.
- Public assistance—numbers in receipt of relief.
- Cost of living index number.
- Board of trade index number (wholesale prices).
- Trade disputes—general figures, summaries.
- Trade unions ,, ,,
- Workmen's compensation ,, ,,

Local Unemployment Index. (Monthly.)

Details and summaries of the cause of unemployment—men, women, juveniles—each of 747 representative towns—county areas.

Report of the Ministry of Labour. (Annual.)

- Details and summaries in regard to the work of employment exchanges, giving figures in respect of men, women and juveniles. analysed under industries, &c.
- Cost of living index number—monthly figures.
- General statistics as to time rates, wages, disputes, trade unions, relief of unemployment, unemployment insurance—unemployment fund.

Unemployment fund.

Receipts and Payments of the Fund, general statistics—report of Comptroller and Auditor-General.

Abstract of Labour Statistics. (Annual.)

A general summary of all items relating to labour, employment and unemployment, workpeople, &c., including the following:—

Industrial accidents Industrial diseases	fatal, non-fatal, each industry—cause.

Workmen's compensation—general summary—claims, fatal—disablement, &c.
Building societies—number, membership, advances, capital, assets.
Number of employees in certain industries—coal mining, fishing, agriculture, railways, shipping, &c.
Employment and unemployment—general statistics and summaries analysed under industries, sex, &c.
Conditions of employment—hours of labour, minimum time rates, rates of wages, &c.
Occupations—industry, sex, &c.
Poor relief—general statistics, numbers, &c., institutional, domiciliary, casual relief.
Trade unions—number, membership, industry, funds.
Trade disputes—numbers involved, cause, result, duration, time lost, industry.
Employers' associations—number, each industry, &c.
Friendly societies—number, membership, funds.
Profit sharing and co-partnership—number of schemes, employees participating, bonus paid, industries.
Industrial assurance—number of societies, assurances, income, expenditure.
Population—number, sex—age.
Migration—general figures.
National Health and Pensions Insurance—general figures, funds, receipts, expenditure.
Old Age Pensions—payments, number, sex, amount paid.

Cost of living index number Wholesale prices index number	Monthly changes.

National Debt Commissioners.

(*a*) Sinking Funds.

Accounts of the Commissioners of the management of the Old and New Sinking Funds.

(*b*) Local Loans Fund.

Receipts and payments in regard to the Capital and Income of the Local Loans Fund, together with general summaries of the transactions of the fund.

National Savings Committee.

Annual Report.

Issues.

- Repayments.
- Principal and interest due to holders.
- Conversion.
- Expenditure—suitably analysed under staff, local committees, associations, &c.
- Loans to Local Authorities (made through the Public Works Loan Commissioners)—number, amount.

Public Works Loan Board.

- Annual Report.
  - Income and expenditure of the Board, suitably sub-divided.
  - Particulars of Loans—
    - made to local authorities (analysed under each type)
    - analysed under the various Acts of Parliament
    - rates of interest
    - periods
    - methods of repayment
    - repaid
    - outstanding
    - in arrear
    - refused.

Queen Anne's Bounty Office.

- Annual Report of the Governors.
  - Income and expenditure
  - Details of securities held
  - Details of grants made
  - Tithe rentcharge
    - Amounts due
    - Amounts collected.

Board of Trade.

- Board of Trade Journal. (Weekly.)
  - Industrial accidents
  - Bankers clearing houses—total, daily averages.
  - Imports and exports of bullion.
  - Preliminary reports of census of production.
  - Imports and exports—all articles.
  - Registered companies—new issues of capital, &c.
  - General statistics as to output, hours of labour, &c., in respect of coal, cotton, electricity, engineering, railways.
  - General statistics, summaries, &c., as to imports and exports—all articles.
  - Freight traffic—railways—tonnage, number, receipts, shipping.
  - Index number of industrial production.

Wholesale prices index number.
Migration—figures and summaries.
Employment and unemployment—general figures and summaries.

Return relating to all authorised Gas Undertakings in Great Britain. (Annual.)

Part 1 Particulars relating to the manufacture and supply of gas. Production, type, quantity, calorific value; coal, coke, oil utilised, residual—coke, breeze, tar, &c. Consumption number, gas sold, lengths of mains, general details and summaries.

Part 2 Finance and Prices.
Each company—capital, dividend authorised and paid, paid up capital, loan capital—authorised, issued, rates of interest; receipts, expenditure, surplus balances. Each local authority—loans authorised, raised, outstanding, repaid, interest payable, sinking funds, revenue and expenditure, appropriation of net profit.

Census of Production. (1907, 1912, 1924, 1930).
Trades within the undermentioned groups of Industries.
(1) Textiles, leather, clothing.
(2) Iron and steel, engineering, shipbuilding, vehicles, non-ferrous metals.
(3) Food, drink, tobacco, chemicals, paper, printing and stationery.

Production—Output, gross and net values, per person employed.
    Quantities and values of goods made.
    Comparisons with results of previous census.
Prices—Goods or work.
Exports—Quantities.
    Proportion of production.
Employees—Operatives.
    Administrative.
    Outworkers.
Wages—Amounts.
    Proportion of net output.
Power—Mechanical power:
    in use
    in reserve
Fuel—Consumption for power for other purposes.

Ministry of Transport. (Annual.)
Report of the Administration of the Road Fund.

- Vehicle and driving licences—general summaries.
- Receipts and payments.
- Motor tax account—receipts and payments.
- Particulars and summaries as to accidents.
- Maintenance, &c., of roads—Class 1—Class 2—unclassified—mileages, expenditure per mile, type of authority, loan charges, &c.
- Construction and improvement of roads and bridges—cost, progress made, programmes, cost to Road Fund.
- Road Fund.
    - Receipts and payments—report of the Comptroller and Auditor-General.

The Treasury. (Annual.)

(*a*) Income and expenditure of the United Kingdom.
- Balances brought forward.
- Receipts and payments for the year—suitably summarised.
- Balances carried forward.

(*b*) Finance Accounts of the United Kingdom.
- Income and expenditure—detailed analysis under each head.
- National Debt—issues and reduction—analysed under types, funded, unfunded, &c.
- Loans to Foreign, Dominion, &c., Governments—amounts outstanding and interest received.
- Loans guaranteed by the Government—details of amounts outstanding.

(*c*) Financial Statement. (Presented with the opening of the Budget).
- Income and expenditure—detailed analysis, under each head showing estimated and actual.
- Rates of the various duties, &c., leviable.
- Alterations in taxation—estimated loss or gain to the Exchequer.
- National Debt—general particulars.

(*d*) National Debt.
- Particulars analysed in detail over a number of years of:—
    - Liabilities, assets and cash balances.
    - Issues and cash value of issues.
- Other transactions in regard to the management of the National Debt—conversion, &c.

(*e*) Civil Estimates.

Showing, in detailed analysis, particulars under the undermentioned headings, together with relevant statistics and an index.
- General Memorandum.
- Class 1. Central Government and Finance.

2. Foreign and Imperial.
3. Home Department.
4. Education.
5. Health, labour and insurance.
6. Trade and industry.
7. Common services.
8. Pensions, &c.
9. Exchequer contributions to local authorities.

(*f*) Public Social Services.

A general return giving particulars of expenditure on social services and showing the number of persons benefiting.

(*g*) Public Departments.

Civil Service and Revenue Departments and the Fighting Services—gross and net costs—votes, receipts, &c.

Bank of England.

Monthly Statistical Summary.

From January, 1946, the contents of this summary form part of the 'Monthly Digest of Statistics' issued by the Central Statistical Office.

1. Issue and Banking Departments.
2. Currency.
3. London Clearing Banks.
4. Exchequer Transactions.
5. Floating Debt.
6. Savings.
7. Wholesale Prices
8. Cost of Living.
9. Retail Trade.
10. Security Prices.
    British Government Securities.
    Investors Chronicle Index.
11. Profits.
12. Wages.
13. Unemployment.
14. Trade Disputes.

London County Council.

London Statistics.

Detailed statistics and summaries in regard to the local government and other aspects of London, e.g.—

Authorities and areas
Population and franchise
Vital and meteorological statistics
Details in regard to labour and unemployment

Public assistance, charities and old age pensions
Public health, hospitals and drainage
Housing
Baths and washhouses
Lunacy and mental deficiency
Open spaces and parks
Libraries and museums
Police, administration of justice, fire brigade and ambulance services
Education, higher and elementary
Highways
Supply of water, gas and electricity
Rates levied, valuation
Income and expenditure
Capital expenditure and loan debt.

Institute of Municipal Treasurers and Accountants. (Annual.)

Return of rate collection

- Authorities with a population of 20,000 or over
- Rateable values
- Rate in the £
- Number of Assessments
- Amount to be collected
- Percentage actually collected
- Percentage uncollected, irrecoverables and allowances
- Particulars of compounded property
- Discounts for prompt payments.

Return of Rates Levied in Various Towns (*The Borough Treasurer, Preston*). (Annual.)

Comparative information in regard to

84 County Boroughs.
157 Municipal Boroughs.
84 Urban Districts.
28 Metropolitan Boroughs.

A. General information, e.g. population, rateable value, Product of 1^d. Rate, &c.

B. Rate poundages of various services.
Rate aid to Trading undertakings.
Rate relief received from Trading undertakings.
Balances.
Exchequer Grants.
Rates levied.

C. Transfers to and from Trading undertakings.

D. Charges for gas.
water.
E. Charges for electricity.

Return of the Loan Indebtedness of County Councils (*The County Accountants Society*). (Annual.)

Comparative information as indicated in the title is given, suitably analysed and summarised.

Rates Levied—Counties (*The County Accountants Society*). (Annual.)

General information as to the rates levied, &c., by County Councils.

Return of the Loan Indebtedness of County Borough Councils (*The Borough Treasurer, South Shields*).

Comparative information as indicated in the title is given, suitably analysed and summarised.

Return of Rates levied in London Boroughs (*The Borough Treasurer, Southwark*). (Annual.)

Rates levied in respect of each service in respect of London Boroughs and Precepting Authorities.

Amounts of General Exchequer and additional grants and additions or reductions in respect of adjustments of grant.

Rates levied per Head of Population (*The Borough Treasurer, Chesterfield*).

Rates levied per head of the population in various towns, analysed under the headings of services provided.

Miscellaneous Statistical Return (*The County Accountants' Society*).

County.
Acreage.
Population—Boroughs, Urban Districts, Rural Districts, Total.
Rateable Value—Industrial and Freight Transport Hereditaments, Other Hereditaments, Crown Properties, Total.
Rateable Value per head of population.
Estimated Product of Penny Rate.
Ratio of Weighted to Unweighted Population (Block Grant).
Mileage of County Roads—Classes 1, 2, Unclassified, Total.
Mileage of Trunk Roads.
Children on School Registers (Primary and Secondary)—Number, Product of Penny Rate per number of Children on Registers.
Relief—Persons in Receipt—Industrial, Domiciliary, Total—per 10,000 population—Average Weekly Cost of Out Relief per Person.
Mental Hospitals—Number of Patients.
Police—Authorised Establishment—Average daily strength—population per Constable.

LIST OF SYMBOLS

a .. arithmetic average.
c .. class-interval.
d .. deviation from an average.
d_m .. deviations from the median.
d_z .. deviations from the mode.
d_x .. deviations from an assumed or approximate average.
δ .. average deviation.
δ_m .. average deviations from the median.
δ_z .. average deviations from the mode.
f .. the frequency or number of items in a class.
f_1 .. number of items in the class below that in which the mode is situated.
f_2 .. number of items in the class above that in which the mode is situated.
i .. position or order of the median in the class in which it is situated.
j_1 .. 1st coefficient of skewness.
j_2 .. 2nd coefficient of skewness.
j_3 .. 3rd coefficient of skewness.
L .. lower limit of a class.
M .. median.
m .. value of items in a class or series.
n .. number of items.
P .. price or value used in connection with index numbers.
Q_1 .. lower quartile.
Q_3 .. upper quartile.
r .. coefficient of correlation (Karl Pearson).
Σ .. the sum of.
σ .. standard deviation.
v .. coefficient of variation.
w .. weight applied.
Z .. mode.

INDEX

A PAGE

Accuracy and Approximation .. 27
    Error .. 27
    Absolute Error .. 27
    Relative Error .. 27
    Percentage Error .. 28
    Unbiased Error .. 28
    Biased Error .. 29
    Error in Totals .. 28
    Error in Averages .. 29
    Inexact Numbers—
        Addition .. 31
        Subtraction .. 31
        Division .. 31
        Multiplication .. 31
Agriculture and Fisheries, Ministry of .. 160
Applied Statistics .. 17
Arithmetic Average (Mean) .. 63
    Advantages .. 63
    Disadvantages .. 63
    Weighted .. 64, 76
    Moving .. 65
    Descriptive .. 66
    Progressive .. 66
Averages .. 62
    Characteristics .. 62
    Advantages of .. 62
    Disadvantages of .. 63
    Arithmetic Average (Mean) .. 63
    Weighted Arithmetic Average .. 64, 76
    Moving Average .. 65
    Descriptive Arithmetic Average .. 66
    Progressive Average .. 66
    Geometric Mean .. 67, 77
    Harmonic Mean .. 67, 77
    Percentages .. 68
    'Per mille' .. 69
    Ratios .. 70
    Mode (Norm) .. 70, 76
    Median .. 71, 77
    Quartiles .. 73, 77
    Deciles .. 73
    Percentiles .. 74
    Cautions in use .. 78
Average Deviation .. 107, 116

B

'Bankers' Magazine' Investment Number .. 104
Bar Diagrams .. 34
Binomial Expansion .. 131
Birth Rate .. 146
Block Diagrams .. 34
Board of Control (Lunacy and Mental Deficiency) .. 161
Bowley, Prof., Wage Index .. 105
Business Activity, Index of .. 103

C

PAGE

Cartograms .. 37
Catastrophic Movements .. 79
Cautions .. 15
Central Statistical Office .. 159
'Chesterfield' Return .. 174
Circles .. 34
Class .. 25
Classification—
  Objects .. 22
  General Considerations .. 22
  Rules and Aids to Clarity .. 23
  Mechanical Aids .. 157
Colonial Office .. 161
Correlation .. 118
  Nature of .. 118
  Investigation of .. 119
    Galton Graph .. 121
    Regression .. 122, 125
    Coefficient of Correlation .. 122
    General Trend .. 127
    Short Term Fluctuations .. 127
Cost of Living Index Number .. 97
County Accountants Society .. 174
Cumulative Frequency Tables .. 26
Cyclical Fluctuations .. 79

D

Data—
  Preliminary Considerations .. 19
  Sources .. 19
  Methods .. 19
  Questionnaires .. 20
  Original .. 22
  Derivative .. 22
  Descriptive .. 22
  Scientific .. 22
Death Rate (Crude) .. 146
Death Rate (Corrected) .. 147
Deciles .. 73
  Ascertainment of .. 73
Definition .. 13
Diagrammatic Presentation—
  Advantages .. 33
  Disadvantages .. 33
Digest of Statistics, Monthly .. 159
Dispersion (Deviation) .. 106
  Graphic Methods of illustration .. 106
  Moments of .. 107
  Measures (Absolute and Relative) .. 107
    The Range .. 107
    Average Deviation .. 107, 116
    Standard Deviation .. 109, 116
    The Modulus .. 111
    Quartile Measure .. 111, 117
    Skewness .. 112, 117
      First Measure .. 114
      Second Measure .. 114
      Third Measure .. 115

E

PAGE
'Economist' Index of Business Activity .. 103
  Prices Index Number .. 101
Education, Ministry of .. 161
Electricity Commission .. 162
Equated Values .. 45
Error .. 27
  Absolute .. 27
  Relative .. 27
  Percentage .. 28
  Unbiased .. 28
  Biased .. 29
  In Totals .. 28
  In Averages .. 29

F

'Financial Times' Index Number .. 102
Fluctuations .. 79, 82
  Short Term .. 81, 82
Frequency Curves .. 50
  Construction of .. 50
  Frequency Polygon .. 51
  'Smoothed' Curve .. 51
  'Normal Curve of Error' .. 52
  'Smoothing' .. 52
  Skewness .. 53
  Ogives .. 53
Frequency Tables .. 25
  Preparation of .. 26
  Cumulative Frequency Tables .. 26

G

Galton Graph .. 121
Gantt Chart .. 59
General Register Office .. 163
Geometric Mean .. 67, 77
  Advantages .. 67
  Disadvantages .. 67
Graphs .. 39
  Characteristics .. 39
  Construction .. 41
  Logarithmic .. 46
  Fequency Curves .. 50
  Lorenz Curves .. 55
  Galton Graphs .. 121
  Z Chart .. 58
  Gantt Chart .. 59
Group .. 25
Guide to Current Official Statistics .. 159

H

Harmonic Mean .. 67, 77
  Advantages .. 67
  Disadvantages .. 67
Health, Ministry of .. 164
Histogram .. 53
Historigram .. 53
Home Office .. 166

I

Index Numbers .. 89
  Nature .. 89
  Factors .. 89

INDEX NUMBERS (*continued*) PAGE
  Cautions 89
  Construction 92
  Simple Index Numbers 92
  Weighted Index Numbers 93
  Fixed Base Method 92
  Chain Base Method 95
  Published Index Numbers 97
INDUSTRIAL PRODUCTION, INDEX OF 99
'INERTIA OF LARGE NUMBERS' 134
INFANT MORTALITY RATE 146
INLAND REVENUE, BOARD OF 166
INSTITUTE OF MUNICIPAL TREASURERS AND ACCOUNTANTS 173
INVESTMENT INDEX NUMBER 104
'INVESTORS CHRONICLE' INDEX NUMBER 104

L

LABOUR, MINISTRY OF 167
  Cost of Living Index Number 97
LIMITATIONS 14
LINE CHARTS, see GRAPHS 39
LOGARITHMIC CURVE 46
LONDON COUNTY COUNCIL 172
LORENZ CURVE 55

M

MARRIAGE RATE 146
MEANS OF SAMPLES, DISTRIBUTION 139
MECHANICAL AIDS 157
MEDIAN 71, 77
  Ascertainment 71
  Interpolation 72
  Advantages 72
  Disadvantages 73
MEDICAL OFFICER, STATISTICS FOR 146
MODE (NORM) 70, 76
  Ascertainment 70
  Formula 71
  Advantages 71
  Disadvantages 71
MODULUS 111
MOMENTS OF DISPERSION 107
MONTHLY DIGEST OF STATISTICS 159
MUNICIPAL STATISTICS 140
  For Ratepayers 140
  For the Treasurer 144, 152
  For the Medical Officer 146
  For the Surveyor 151, 153
  Investigations, &c. 154

N

NATIONAL DEBT COMMISSIONERS 168
NATIONAL SAVINGS COMMITTEE 168
NATURE 13
NORM, *See* the 'Mode' 70, 76
'NORMAL FREQUENCY CURVE' 129, 132
NUMBERS INEXACT—
  Mathematical Treatment—
    Addition 31
    Subtraction 31
    Division 31
    Multiplication 31

O

PAGE

- Ogives .. 53

P

- Pearson (Coefficient of Correlation) .. 122
- Percentages .. 68
  - Advantages .. 68
  - Disadvantages .. 68
- Percentiles .. 74
- ‘Per Mille’ .. 69
- Pictograms .. 33
- Pie Charts .. 34
- Polygon, Frequency .. 51
- Population, Estimation .. 149
- Preston Rates Return .. 173
- Prices Index .. 100
- ‘Probability’ .. 129
  - Non-Measurable Characteristics .. 136
  - Measureable Characteristics .. 137
- Profits Index .. 103
- ‘Proportional Numbers,’ *see* Index Numbers .. 89
- Public Works Loan Board .. 169
- Published Statistics .. 159
  - Guide to Current Official Statistics .. 159
  - Monthly Digest of Statistics .. 159
  - Statistical Abstract of the United Kingdom .. 160
  - Ministry of Agriculture and Fisheries .. 160
  - Colonial Office .. 161
  - Board of Control (Lunacy and Mental Deficiency) .. 161
  - Ministry of Education .. 161
  - Electricity Commission .. 162
  - General Register Office .. 163
  - Ministry of Health .. 164
  - Home Office .. 166
  - Board of Inland Revenue .. 166
  - Ministry of Labour .. 167
  - National Debt Commissioners .. 168
  - National Savings Committee .. 168
  - Public Works Loan Board .. 169
  - Queen Anne’s Bounty Office .. 169
  - Board of Trade .. 169
  - Ministry of Transport .. 170
  - The Treasury .. 171
  - London County Council .. 172
  - Institute of Municipal Treasurers and Accountants .. 173
  - ‘Preston’ Rates Return .. 173
  - County Accountants’ Society .. 174
  - ‘South Shields’ Return .. 174
  - ‘Southwark’ Return .. 174
  - ‘Chesterfield’ Return .. 174

Q

- Quartiles .. 73, 77
- Quartile Measure of Dispersion .. 111, 117
- Queen Anne’s Bounty Office .. 169
- Questionnaires, General Considerations .. 20
  - Characteristics .. 20

R

- Range of Data .. 107
- Ratepayers, Statistics for .. 140
- Ratios .. 70

PAGE
RECTANGLES 34
REGRESSION 122,125
REPRODUCTION RATE 147

S

SAMPLING 24
  Attributes 24
  Methods 25
  Distribution of Means of Samples 137
SAUERBECK, INDEX OF AVERAGE PRICES 100
'SCIENCE OF AVERAGES' 13
'SCIENCE OF ESTIMATES AND PROBABILITIES' 13
SEASONAL VARIATIONS 79
SERIES 25
  Continuous 25
  Discrete 26
SKEWNESS 112, 117
  Measurement 112
'SMOOTHING' 51
  Justification for 51
  Extent 51
'SOUTH SHIELDS' RETURN 174
'SOUTHWARK' RETURN 174
SQUARES 34
STAMP, LORD, INDEX OF AGGREGATE PROFITS 103
STANDARD DEVIATION 109, 116
Statist INDEX NUMBER 100
STATISTICAL DEPARTMENT 18
STATISTICAL METHODS 17
STATISTICAL OPERATIONS 17
STATISTICAL REGULARITY, LAW OF 134
SURVEYOR, STATISTICS FOR 151, 153
STATISTICAL SIGNIFICANCE 135
STATISTICAL ABSTRACT OF THE UNITED KINGDOM 160

T

TABULATION, OBJECTS 22
  General Considerations 23
  Rules and Aids to Clarity 23
  Mechanical Aids 157
'TIME LAG' 128
TRADE, BOARD OF 169
  Index Number (Prices) 101
  Index of Industrial Production 99
TRANSPORT, MINISTRY OF 170
TREASURER, STATISTICS FOR 144, 152
TREASURY 171
TREND 79
  Secular 79
  Periodic 79
  Investigation of 80, 83

V

VALUE 13
VARIATION, COEFFICIENT OF 111
VARIANCE 109
VARIATION, RATIO OF 126

W

WAGE INDEX 105
'WEIGHTING' 64, 93

Z

Z-CHART 58